

Effects of [N II] and H α line blending on the *WFIRST* Galaxy redshift survey

Daniel Martens,¹ Xiao Fang^{1,2*}, M. A. Troxel,¹ Joe DeRose,^{3,4}
Christopher M. Hirata,¹ Risa H. Wechsler^{3,4,5} and Yun Wang⁶

¹Center for Cosmology and AstroParticle Physics, Department of Physics, The Ohio State University, 191 W Woodruff Ave, Columbus, OH 43210, USA

²Department of Astronomy and Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85719, USA

³Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

⁴Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

⁵SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

⁶IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

Accepted 2019 February 6. Received 2019 January 19; in original form 2018 April 21

ABSTRACT

The Wide Field Infrared Survey Telescope (*WFIRST*) will conduct a galaxy redshift survey using the H α emission line primarily for spectroscopic redshift determination. Due to the modest spectroscopic resolution of the grism, the H α and the neighbouring [N II] lines are blended, leading to a redshift bias that depends on the [N II]/H α ratio, which is correlated with a galaxy's metallicity, hence mass and ultimately environment. We investigate how this bias propagates into the galaxy clustering and cosmological parameters obtained from the *WFIRST*. Using simulation, we explore the effect of line blending on redshift-space distortion and baryon acoustic oscillation (BAO) measurements. We measure the BAO parameters α_{\parallel} , α_{\perp} , the logarithmic growth factor f_v , and calculate their errors based on the correlations between the line ratio and large-scale structure. We find $\Delta\alpha_{\parallel} = 0.31 \pm 0.23$ per cent (0.26 ± 0.17 per cent), $\Delta\alpha_{\perp} = -0.10 \pm 0.10$ per cent (-0.12 ± 0.11 per cent), and $\Delta f_v = 0.17 \pm 0.33$ per cent (-0.20 ± 0.30 per cent) for redshift 1.355–1.994 (0.700–1.345), which use approximately 18 per cent, 9 per cent, and 7 per cent of the systematic error budget in a root-sum-square sense. These errors may already be tolerable but further mitigations are discussed. Biases due to the environment-independent redshift error can be mitigated by measuring the redshift error probability distribution function. High-spectral-resolution reobservation of a few thousand galaxies would be required (if by direct approach) to reduce them to below 25 per cent of the error budget. Finally, we outline the next steps to improve the modelling of [N II]-induced blending biases and their interaction with other redshift error sources.

Key words: line: identification – surveys – cosmology: observations.

1 INTRODUCTION

Observations of high-redshift supernovae in the 1990s provided the first direct evidence for an acceleration in the expansion rate of the Universe (Riess et al. 1998; Perlmutter et al. 1999). Whatever field or particle is responsible for this surprising acceleration has been dubbed ‘dark energy’, and one of the major observational programs in modern cosmology is to measure its properties. It is of particular interest to determine whether the dark energy is consistent with a cosmological constant; whether it requires new dynamical degrees

of freedom, or whether cosmic acceleration arises instead from a modification to the laws of gravity on large scales.

In order to measure the properties of dark energy, cosmologists employ a variety of techniques. Observations of supernovae provide information on the expansion rate of the Universe for different redshifts, probing the effects of dark energy throughout cosmic history. Weak lensing surveys probe the matter distribution, allowing measurements of clustering at various redshifts. Galaxy clusters are the most massive collapsed objects produced by cosmological structure formation, and can be traced using a wide range of observables (the visible galaxy content, the hot gas via X-ray emission or the Sunyaev–Zel’dovich effect, and weak lensing). Galaxy redshift surveys – the subject of this paper – can trace

* E-mail: xfang@email.arizona.edu

cosmic structures in three dimensions, although their cosmological interpretation requires accurate modelling of the relation between the visible galaxies and the mostly unseen matter.

The size of redshift surveys has been steadily increasing in tandem with technological improvements. A sample of over 200 000 galaxies was investigated in the 2dF Galaxy Redshift Survey (2dFGRS), constraining cosmological parameters within specific cosmological models (Cole et al. 2005). This was followed by the 6dF Galaxy Survey (Jones et al. 2009). A spectroscopic analysis of over 54 000 luminous red galaxies using the Sloan Digital Sky Survey (SDSS) found evidence for baryon acoustic oscillations (BAOs), and provided additional constraints on the cosmological parameters (Eisenstein et al. 2005). Further analysis has been done combining SDSS with 2dFGRS, as well as analysing data provided by SDSS DR7 (Percival et al. 2007, 2010). More recently, the WiggleZ Dark Energy Survey has been used to measure the BAO peak at different redshifts (Blake et al. 2011), while others have analysed the distance to these redshifts (Xu et al. 2012). The SDSS-III/BOSS project, which included an upgraded spectrograph with enhanced red sensitivity, collected spectra of over 2.4 million galaxies (Alam et al. 2015). The redshift range of SDSS-III is extended by SDSS-IV (eBOSS) which is currently observing (Dawson et al. 2016; Blanton et al. 2017).

This progress is expected to continue: the Taipan survey will look at low-redshift galaxies, over half the sky (da Cunha et al. 2017), and the Dark Energy Spectroscopic Instrument (DESI) will conduct a comprehensive spectroscopic survey of galaxies and quasars over the Northern sky (DESI Collaboration et al. 2016a,b). A substantially deeper survey is planned by the Prime Focus Spectrograph (PFS) (Tamura et al. 2016), which will extend ground-based spectroscopic coverage out to 1.26 μm , and the 4m Multi Object Spectroscopic Telescope (4MOST Depagne 2015) project will conduct optical spectroscopy in the south. The *Euclid* mission will conduct a space-based near-infrared grism survey with portions of its footprint in both hemispheres (Laureijs et al. 2011).

The Wide Field InfraRed Survey Telescope (*WFIRST*) will be a 2.4 m space telescope that carries out a wide range of investigations in cosmology, exoplanets, and other areas of astrophysics (Dressler et al. 2012; Green et al. 2012; Spergel et al. 2015) following its launch in the mid-2020's. *WFIRST* will carry a Wide Field Instrument (WFI; capable of imaging and slitless spectroscopy) and a coronagraph. The galaxy redshift survey program on *WFIRST* will use the grism to observe emission lines in the 1.00–1.93 μm bandpass¹ and obtain redshifts for 1.8×10^6 galaxies per month of observations. The principal tracer of large-scale structure will be the $\text{H}\alpha$ emission line (at 6565 \AA), which is visible in *WFIRST* out to a maximum redshift of $z = 1.94$; at higher redshift, other emission lines will be used, most notably the [O III] doublet (4960 and 5008 \AA), which is visible out to $z = 2.85$.

The grism spectroscopy technique has the advantage of simplicity (the grism occupies one slot on the *WFIRST* filter wheel, with no additional moving parts); it provides enormous multiplexing, and it does not require that targets be selected in advance (thus providing operational flexibility, and avoiding selection biases that are ‘baked in’ to traditional redshift surveys at the time of target selection). However, it does have drawbacks. One is that without a slit, each pixel is exposed to the full dispersed sky scene rather than

only the targeted galaxy, which leads to higher backgrounds and confusion from other sources. Grism spectroscopy also has some constraints: *WFIRST* requires a wide-field grism in a converging beam; it was a significant design challenge for all field positions and all wavelengths to focus simultaneously, and solutions are only available at moderate spectral resolution (Gong et al. 2016). At this resolution ($R \sim 690$ per two-pixel element at $\lambda = 1.5 \mu\text{m}$), the $\text{H}\alpha$ line is partially blended with the neighbouring [N II] lines. This is a similar situation to the WFC3 Infrared Spectroscopic Parallel (WISP) Survey, which used an even lower resolution grism on the *Hubble Space Telescope* where $\text{H}\alpha + [\text{N II}]$ was completely blended (Atek et al. 2010). The results of the WISP survey therefore have been used to make predictions on the effectiveness of future surveys (Colbert et al. 2013; Mehta et al. 2015a; Merson et al. 2018), particularly through the discussion of the changes in the luminosity function of galaxies when line-blending issues are present. In the past, a correction factor of 0.71 has been applied to these $\text{H}\alpha$ luminosities of galaxies to account for [N II] contamination (Mehta et al. 2015b).² Further analyses of the effect of blending on the luminosity function, as well as an empirical parametrization of the [N II]/ $\text{H}\alpha$ flux ratio as a function of galaxy properties and redshift, were completed by Faisst et al. (2018).

A separate concern arising from the line-blending effect is the change in the observed redshift. Since the [N II] lines are asymmetric in emission strength, they will shift the ‘observed’ redshift z_{obs} (assuming that the line centroid is at the $\text{H}\alpha$ wavelength in the rest frame), to be redder than the true redshift, z_{true} (Faisst et al. 2018). Since the line ratio is determined by H II region physics and depends on metallicity, this effect is not the same for all galaxies: it is not removable by subtracting a mean bias. Furthermore, since metallicity is correlated with galaxy mass and hence with large-scale structure, $\text{H}\alpha + [\text{N II}]$ line blending leads to a redshift bias that is correlated with large-scale structure and could have highly non-trivial effects on the inferred cosmological parameters.

The purpose of this paper is to quantify the extent of this observational problem, understand how it will affect the measured galaxy correlation function if left unmitigated, and discuss potential mitigation strategies to the extent that they will be necessary. We will analyse a sample of $>10^8$ galaxies using mock data from the BUZZARD-V1.1 simulation. We compute the correlation function and perform BAO and RSD fits with both the true redshift catalogue and the observed redshift catalogue and assess the differences. We also create a ‘shuffled’ redshift catalogue, where the redshift errors $\Delta z/(1+z)$ are scrambled, which allows us to test which changes in the clustering properties are due to the correlations of redshift error with large-scale structure, and which are due to the distribution of redshift errors alone.

The outline of the paper is as follows: in Section 2, we discuss the line-blending problem in more detail, and introduce the main questions which this paper will attempt to answer. In Section 3, we describe the general road map to answering these questions, discussing the analysis strategies we will use to dissect the simulation results. We discuss details of our simulation in Section 4, as well as the description of the catalogues. In Section 5, we discuss our fitting strategies and methods, and display our fits for redshift-space

¹This wavelength range was chosen for the System Requirements Review/Mission Definition Review, and is somewhat different from that considered during previous iterations of the *WFIRST* design.

²Other studies have used different methods to account for the [N II] contamination. For example, Sobral et al. (2013) use a relation between [N II]/ $\text{H}\alpha$ ratio and [N II] + $\text{H}\alpha$ line equivalent width as described in Villar et al. (2008) to de-blend their [N II] + $\text{H}\alpha$ fluxes derived from narrow-band observations.

distortion (RSD) and BAO parameters to the data. In Section 6, we discuss our results and how they compare to the requirements for *WFIRST*, with a brief discussion of possible mitigation strategies. We conclude in Section 7.

For this analysis, we use a flat Λ CDM cosmology with parameters of $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\sigma_8 = 0.82$, $h = 0.7$, $\Omega_b = 0.047$, and $n_s = 0.96$. This is consistent with the parameters used in the generation of the BUZZARD-V1.1 simulations. Finally, all line wavelengths are referenced to their vacuum values.

2 OBJECTIVES

All spectroscopic galaxy surveys contain some redshift errors, in the sense that the observed redshift z_{obs} deviates from the true redshift z_{true} . Among existing samples used for large-scale structure analyses, this is most evident for quasars (e.g. Dawson et al. 2016), since the most accessible lines are broad and often asymmetric, and redshift errors of several hundred km s^{-1} are common.

In the case of *WFIRST*, photon noise in the centroid of the emission lines will be the dominant source of scatter in the z_{obs} versus z_{true} relation. The inclusion of statistical noise results in both a decrease in the constraining power of the survey, and the suppression of power at large k_{\parallel} . This noise does not require any additional treatment to remove, since the central values are unchanged. However, in the *WFIRST* survey, blending of the $\text{H}\alpha$ emission line with the neighbouring [N II] doublet (one doublet member is on each side of $\text{H}\alpha$) results in an offset $z_{\text{obs}} > z_{\text{true}}$ for objects of higher [N II]/ $\text{H}\alpha$ ratio (which are likely higher metallicity and hence probably found in denser environments). Other possible errors could involve the emission line strength and the angular size of a galaxy affecting the width of the $z_{\text{obs}} - z_{\text{true}}$ distribution, and this may in turn affect the redshift offset, if it interacts with point spread function (PSF) asymmetry. Both of these properties may be correlated with the galaxy environment. They are, however, beyond the scope of this paper; we plan to revisit PSF asymmetry and other instrument-related issues when the *WFIRST* grism simulation pipeline is at a more advanced stage.

Fig. 1 demonstrates the scale of the line-blending problem. The *WFIRST* grism has a spectral resolving power of

$$\frac{\lambda}{\Delta\lambda} = 461 \left(\frac{\lambda_{\text{obs}}}{1 \mu\text{m}} \right), \quad (1)$$

where λ is the observed wavelength (this is measured for a two-pixel element; an extended galaxy will be bigger and hence have lower spectral resolution).³ It is seen that at this resolution, the $\text{H}\alpha$ line and surrounding [N II] lines are blended, and a fit to a single line will find something close to the centroid of the blended features. If there was a known error probability distribution function (PDF) $P(z_{\text{obs}} - z_{\text{true}} | z_{\text{true}})$ that was both uncorrelated with galaxy environment and valid for every type of galaxy in the sample, then we could incorporate this in the theoretical correlation function, and the only effect of the redshift errors due to line blending would be a reduction in the statistical constraining power in the survey. If, however, the redshift error PDF is either not known or is correlated with galaxy environment, further steps may be needed to maximize accurate redshift reconstruction. Such correlations may be problematic for a

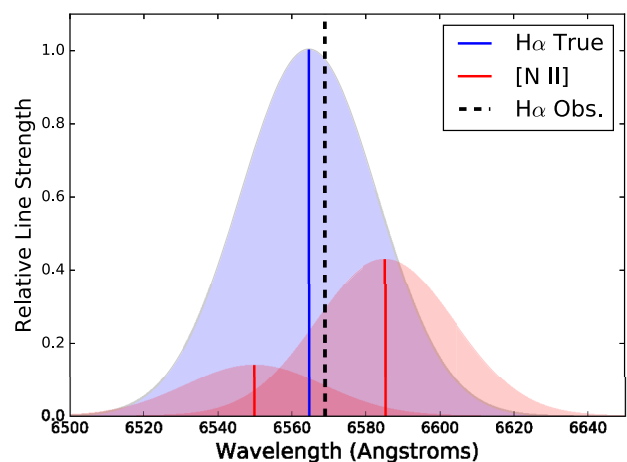


Figure 1. A visual representation of the [N II] lines and $\text{H}\alpha$ line for the line-blending scenario. The difference in strength between the two nitrogen lines is constant for each galaxy with a ratio of approximately 0.32, while the difference between the larger [N II] line and the $\text{H}\alpha$ line varies depending on the metallicity of the galaxy in question. Here, we show the median difference of our sample, with a ratio of [N II]6585/ $\text{H}\alpha$ of 0.427. The Gaussian spreading centred at each line has a standard deviation of σ_{grism} , calculated generally in equation (1). In this example, we have shown the spreading due to a galaxy with radius 4 kpc, at redshift 1.5. The black dotted line at 6569 Å (about 4.3 Å larger than the true $\text{H}\alpha$ wavelength, corresponding to an increase in redshift of about 6.6×10^{-4}) is the resulting observed line, given that the constituent lines cannot be resolved.

survey even if the systematic redshift error is small compared to the statistical errors for one single galaxy.

In this paper, we use the BUZZARD-V1.1 simulation (described in Section 4.1) to address several key questions regarding the impact of the line-blending phenomenon on the *WFIRST* redshift survey:

- (i) If we ignore the effects of line blending, what biases are induced in the BAO and RSD parameters? How does this compare to the corresponding statistical errors on these parameters, or the errors required by the *WFIRST* Science Requirements Document (SRD)?
- (ii) Do we need to mitigate biases caused by correlations between galaxy environment and redshift offsets due to line blending?
- (iii) If we determine that the problems are significant enough to require some level of mitigation, what type of mitigations are necessary?

The *WFIRST* SRD defines required performance based on a Reference Survey of 0.70 yr,⁴ with 1σ statistical errors of 0.70 per cent on the transverse BAO distance scale (α_{\perp}); 1.28 per cent on the radial BAO distance scale (α_{\parallel}); and 1.28 per cent on the rate of growth of structure measured from RSD (f_v). Observational systematic errors are allocated an error of 0.58 times the Reference Survey statistical errors, i.e. 0.41 per cent (α_{\perp}), 0.74 per cent (α_{\parallel}), and 0.74 per cent (f_v).⁵ Note that these are *requirements* – it is always desirable to have smaller systematic errors, but if they exceed their

⁴The actual allocations will be determined in the future by the Implementation SWG.

⁵This means that the combination of statistical errors and observational systematic errors would be $\sqrt{1 + 0.58^2} = 1.16$ times larger than the statistical errors alone. Note furthermore that the SRD allows for other sources of error as well.

³See Gehrels, Spergel & WFIRST SDT Project (2015); updated information can be found at the *WFIRST* project website: <https://wfirst.gsfc.nasa.gov/science/WFIRST-Reference-Information.html>

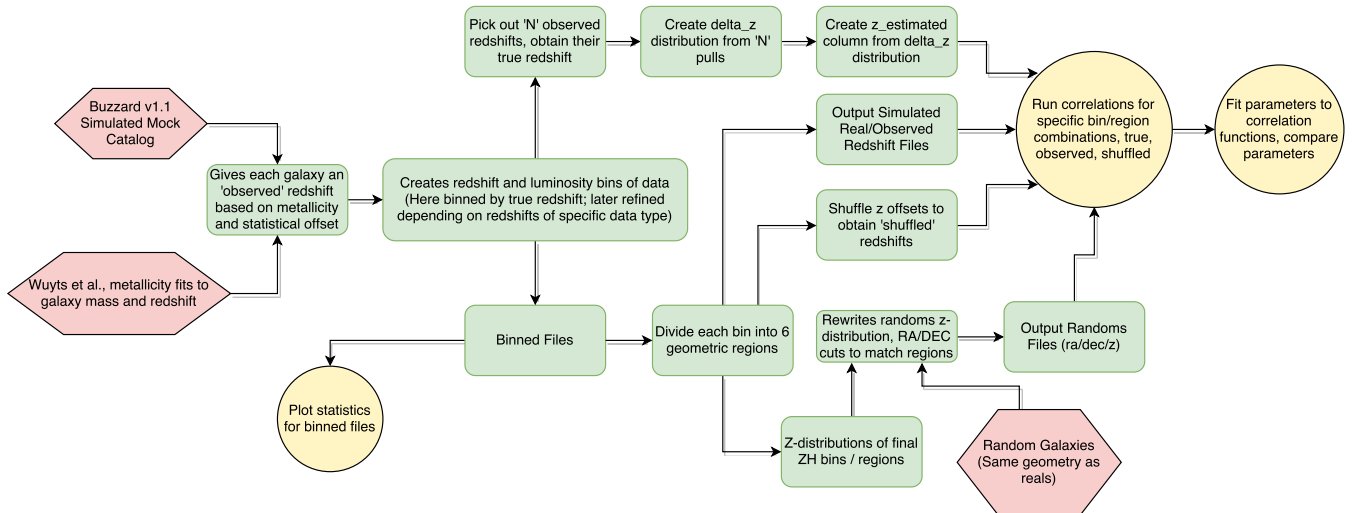


Figure 2. A flow chart describing the simulation pipeline; see Section 3 for details. The red hexagons indicate steps where data are input into the process, and yellow circles indicate steps where we output statistics, correlation functions, or fit parameters. Also included in the flowchart is a method for possible analysis of the ability of a ground-based spectroscopic observation to reproduce the observed catalogue; although it is not implemented in this work, we discuss its potential in Section 6.

allocation they must be mitigated. Finally, note that the $H\alpha + [N II]$ blending is only one type of observational systematic error, and as such should consume only a fraction of the systematic error budget (the exact percentage has not been set; part of the purpose of this paper is to inform this discussion). Other systematic errors include uncertainties in the optical distortion models (both for grism mode and the direct imaging used as a reference); wavelength calibration; detector effects (e.g. flat fielding, cross-talk); and PSF asymmetry (e.g. centroid fitting in the presence of coma). The draft error budget for these in the SRD is 0.23 times the reference statistical errors.

3 CALCULATION OUTLINE

In this section, we explain the specific steps taken throughout the analysis in order to fully understand the detailed effects of line blending. Fig. 2 shows a flowchart of the process, beginning with the BUZZARD-V1.1 mock galaxy catalogue, moving through the calculations we perform, and ending with our parameter estimates. Although more technical details for each process will be expanded later on in Section 4, we take a moment here to give a high-level overview of the entire pipeline.

We begin with the mock catalogue from the BUZZARD-V1.1 simulation. We use a ‘true’ redshift for each galaxy in the catalogue that incorporates peculiar velocity effects, but does not yet contain any line blending or statistical errors. The catalogue and its generation are described in detail in Section 4.1. The catalogue covers one quadrant of the sky (π steradians).

Our first step is to assign to each galaxy a $[N II]/H\alpha$ ratio based upon that galaxy’s redshift and stellar mass. Once each galaxy has a $[N II]/H\alpha$ ratio, we can then calculate the observed redshift z_{obs} for each galaxy, which incorporates the effect of line blending. We also include a statistical offset of the redshift due to photon noise in the centre of the line. These redshifts form the ‘observed redshift’ catalogue.

We next create a separate redshift for each galaxy called the ‘shuffled’ redshift, or z_{shf} . The purpose of this is to provide a redshift catalogue where the distribution of the observed redshift catalogue is accounted for, but removes any correlation between

Table 1. The number of galaxies, in millions, within each bin. This is the sum total of galaxies within each geometric region for the data counts. The galaxy $H\alpha + [N II]$ flux F is measured in units of $10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$. In parenthesis is the label we use to reference each bin throughout the text. Note that the WFIRST flux limit for point sources is $F = 5$ (Spergel et al. 2015), and around 10 for galaxies.

	$F > 8$	$F > 13$	$F > 25$
$0.705 \leq z < 1.345$	–	65.1 (Z2H2)	17.8 (Z2H3)
$1.355 \leq z < 1.994$	59.1 (Z3H1)	24.9 (Z3H2)	5.2 (Z3H3)

Table 2. The median redshift of each bin.

Bin	Median redshift
Z2H2	0.9116
Z2H3	0.8698
Z3H1	1.6136
Z3H2	1.6134
Z3H3	1.6169

the $[N II]/H\alpha$ ratio and the galaxy environment. By later comparing clustering parameters from the ‘shuffled’ redshifts to those from the observed redshifts, we can determine how much of the effect we see is captured in the one-point distribution of redshift errors, and how much depends on environmental correlations.

Next, the galaxies are binned, depending on the galaxy redshift (Z2, Z3) and line flux (H1, H2, H3). The group cuts are described in Table 1. Note that a given galaxy may be in one bin in the true redshift catalogue, but in another bin in the observed redshift catalogue, if the offsets push the observed redshift into another bin. Furthermore, to ease the computational burden, each bin is split into six congruent kite-shaped geometric regions (S1, S2, ... S6) on the sky, each of solid angle $\pi/6$ steradians, as shown in Fig. 3. These sections are counted separately, and then recombined in the analysis of the correlation functions.

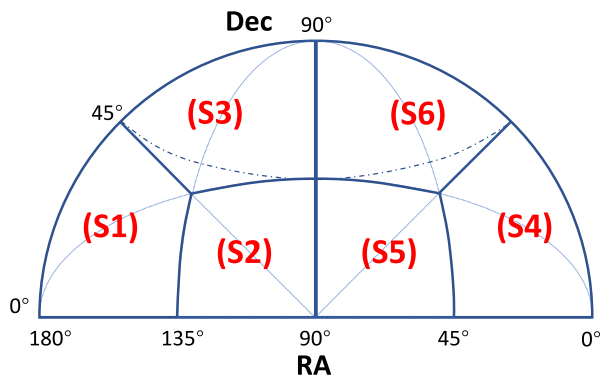


Figure 3. We display the six separate regions on the sky, within which we have independently calculated the correlation function, as discussed in Section 4.3. Note that although we must display a 2D projection, the sectors are congruent on the sphere. (The simulation does not cover the actual *WFIRST* footprint, which is likely to be placed in the Southern hemisphere, but this does not matter in a statistically isotropic universe.)

For each redshift/flux bin (which will be referred to as ‘ZH’ bins) and geometric region, we fit RSD parameters and BAO parameters, and compare the resulting parameter shifts with the *WFIRST* error budget to assess their significance and the possible need for mitigation.

4 SIMULATION

4.1 Catalogue generation

We use the BUZZARD-V1.1 mock galaxy catalogue that we describe briefly here and refer the interested reader to more detailed descriptions in DeRose et al. (2019) and Wechsler et al. (in preparation). BUZZARD-V1.1 is a simulated galaxy catalogue constructed from a set of three nested dark matter-only light-cone simulations which are progressively lower resolution at higher redshifts. The light-cones have volumes of $(1050 h^{-1} \text{Mpc})^3$, $(2600 h^{-1} \text{Mpc})^3$, $(4000 h^{-1} \text{Mpc})^3$, particle masses of $2.7 \times 10^{10} h^{-1} M_{\odot}$, $1.3 \times 10^{11} h^{-1} M_{\odot}$, $4.8 \times 10^{11} h^{-1} M_{\odot}$ and force softenings of $20 h^{-1} \text{kpc}$, $35 h^{-1} \text{kpc}$, $53 h^{-1} \text{kpc}$, respectively. The highest resolution $(1050 h^{-1} \text{Mpc})^3$ simulation is used for $z < 0.34$, the $(2600 h^{-1} \text{Mpc})^3$ for $0.34 < z < 0.9$ and the $(4000 h^{-1} \text{Mpc})^3$ simulation for $0.9 < z < 2.35$. These simulations are run using L-GADGET2, a version of GADGET2 (Springel 2005) modified for memory efficiency with second-order Lagrangian perturbation theory (2LPT) initial conditions created using 2LPTIC (Crocce, Pueblas & Scoccimarro 2006). Light-cones are generated on the fly as the simulations run.

Galaxies are added to the simulation using the ADDGALS algorithm (Wechsler et al., in preparation). Assuming an input luminosity function, this algorithm uses a model for density given absolute magnitude, $p(\delta|M_r, z)$ measured from a subhalo abundance matching (SHAM) model run on a smaller, higher resolution simulation. This model is then applied to the lower resolution light-cone simulations by drawing magnitudes from the assumed luminosity function, drawing densities from $p(\delta|M_r, z)$, and assigning the galaxy to a particle in the light-cone with the correct density. After all rest frame r -band magnitudes are assigned to all galaxies, SEDs are then assigned from SDSS using the $\delta_g - M_r$ –SED relation from SDSS (Cooper et al. 2006), where δ_g is the distance to the fifth-nearest neighbour galaxy. The SEDs are represented by KCORRECT

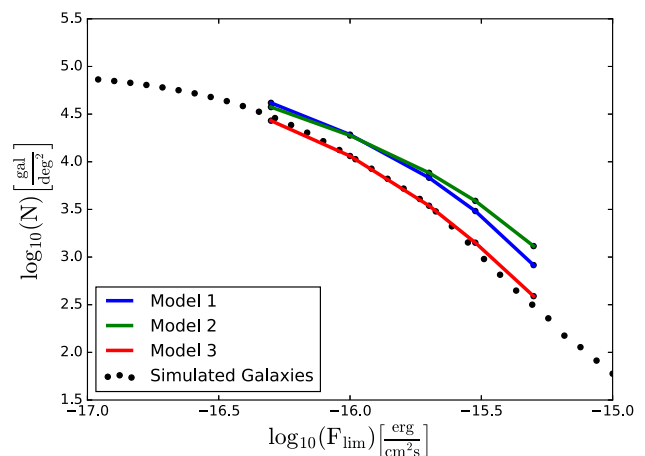


Figure 4. Comparing the luminosity function of the BUZZARD-V1.1 mock catalogue to that predicted by Pozzetti et al. (2016). These are semi-analytic models made by fitting to observed luminosity functions from $H\alpha$ surveys. Specifically, Models 1 and 2 feature a Schechter parametrization, while Model 3 was designed specifically for high-redshift slitless surveys such as *WFIRST* and Euclid, as it was fit directly to luminosity function data. This graphic is from one 54 deg^2 section in the sky; the redshift range is from $z = 0.7$ to $z = 1.5$ [similar to our Z2 bins, but matching the exact range used for the Pozzetti et al. (2016) model].

templates (Blanton et al. 2003) from which line fluxes and stellar masses can be determined. In Fig. 4, we compare the BUZZARD-V1.1 luminosity function to several empirical models based on grism and narrow-band data (Pozzetti et al. 2016). To test the stellar mass function, we take the objects at flux $> 1 \times 10^{-16} \text{ erg cm}^{-2} \text{ s}$, and compare the stellar masses from the catalogue, versus using the flux \rightarrow star formation rate (SFR) \rightarrow stellar mass conversion assuming 1 mag extinction at $H\alpha$ and using the conversion factor for the Chabrier (2003) initial mass function (IMF) and the SFR sequence of Table 1 of Dutton, van den Bosch & Dekel (2010). We obtain that at $z = 1.0 \pm 0.2$ (1.9 ± 0.2), the median stellar mass from the catalogue is 0.23 (0.31) dex above the scaling relation, which is acceptable given the simplicity of the comparison.

4.2 Assignment of the [N II]/ $H\alpha$ ratio

Here we outline our method to generate a [N II]/ $H\alpha$ ratio for each galaxy. The [N II]/ $H\alpha$ ratio is an observable in high-resolution spectra and is a commonly used metallicity indicator (see e.g. Kewley & Dopita 2002; Denicoló, Terlevich & Terlevich 2002; Pettini & Pagel 2004), and so in principle one can use the metallicity to predict the [N II]/ $H\alpha$ ratio. The full picture is more complicated: the Baldwin–Phillips–Terlevich (BPT) diagram of [O III]/ $H\beta$ versus [N II]/ $H\alpha$ (Baldwin, Phillips & Terlevich 1981) has a sequence of star-forming galaxies, ranging from low-metallicity/high-ionization (upper left) to high-metallicity/low-ionization parameter (lower right). The star-forming sequence evolves with redshift, which has been attributed to the N/O ratio varying at fixed metallicity (and instead being more closely correlated with stellar mass; e.g. Masters, Faisst & Capak 2016), or to massive binary stellar populations (e.g. Steidel et al. 2016). We will circumvent this issue for the purpose of this paper using an empirical mass–metallicity (MZ) relation with [N II]/ $H\alpha$ -based metallicities: uncertainties in interpretation of the line ratio cancel out when we predict [N II]/ $H\alpha$ so long as we use the same calibration as in the MZ determination. Creating an accurate representation of this distribution within our simulation – including

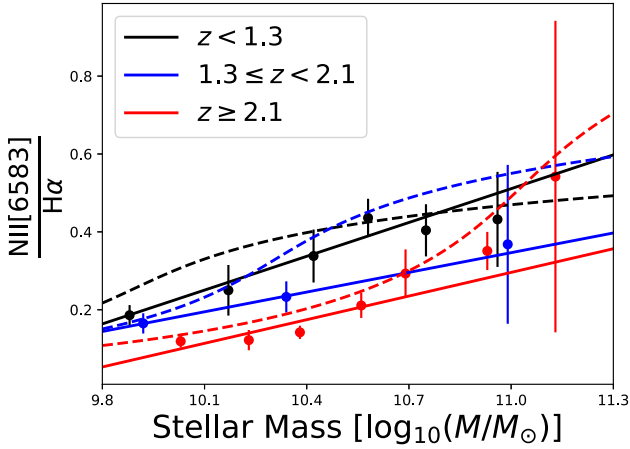


Figure 5. Empirical relationship (solid lines) between $[\text{N II}]/\text{H}\alpha$ ratio and log stellar mass. The data being fitted were obtained from Wuyts et al. (2016). These relationships were used to construct the relative $[\text{N II}]$ line strength within the mock galaxy sample. For comparison, we also plot in dashed lines the fits at redshifts 0.9 (black), 1.5 (blue), 2.3 (red), from equation (3) in Faist et al. (2018). Given the large uncertainties in the data, our fits for redshift larger than 1.3 may or may not underestimate the line ratio. Note that both references adopt the Chabrier (2003) IMF, consistent with the KCORRECT templates in our simulation.

the effects of environment – is critical to the goals of this paper. Furthermore, the $[\text{N II}]/\text{H}\alpha$ ratio is expected to vary as a function of redshift (Kewley et al. 2013; Masters et al. 2014; Steidel et al. 2014; Shapley et al. 2015; Strom et al. 2017; Kashino et al. 2017) similar to the general MZ relation (Savaglio et al. 2005; Maiolino et al. 2008; Lilly et al. 2013).

For each galaxy in the BUZZARD-V1.1 mock catalogue, $[\text{N II}]/\text{H}\alpha$ line strength ratios were assigned based on the stellar mass and redshift of that galaxy; we assume no other environmental trend in the mass–metallicity relationship itself. Specifically, we incorporate only the *mean* trend in the mass–metallicity correlation, without including effects due to the scatter in this relationship. We expect this to comprise the dominant systematic offset (e.g. Lequeux et al. 1979), although including the scatter in the mass–metallicity relationship, which could conceivably be correlated with the galaxy environment (e.g. Cooper et al. 2008), may be an interesting investigation in future work. The dependence of the galaxy metallicity on SFR is accounted for through the correlation between SFR and redshift. This is justified observationally by Wuyts et al. (2016), who used the KMOS near-infrared multi-integral field unit survey to find the $[\text{N II}]/\text{H}\alpha$ ratio for 419 star-forming galaxies, in the redshift range $0.6 < z < 2.7$. They find that there is no significant dependence of the ratio on SFR, given fixed redshift and stellar mass, although several studies have indicated possible connections between the line ratio and the specific SFR (e.g. Mannucci et al. 2010; Lilly et al. 2013; Maier et al. 2015; Masters et al. 2016; Faist et al. 2018). Furthermore, this assumption has support from hydrodynamical simulations; Hirschmann et al. (2017) found that the primary evolutionary trends were based on redshift and stellar mass, in a manner consistent with our model. Other effects such as specifics of their AGN model and ionized-gas hydrogen/electron density were found to have a much smaller impact on the cosmic evolution than the galaxy mass.

We base our fits on the results from Wuyts et al. (2016), who grouped their data into galaxy mass subranges spanning $\log_{10}(M/M_{\odot}) = 9.88\text{--}11.13$, for redshift bins at $z \approx 0.9$, $z \approx 1.5$,

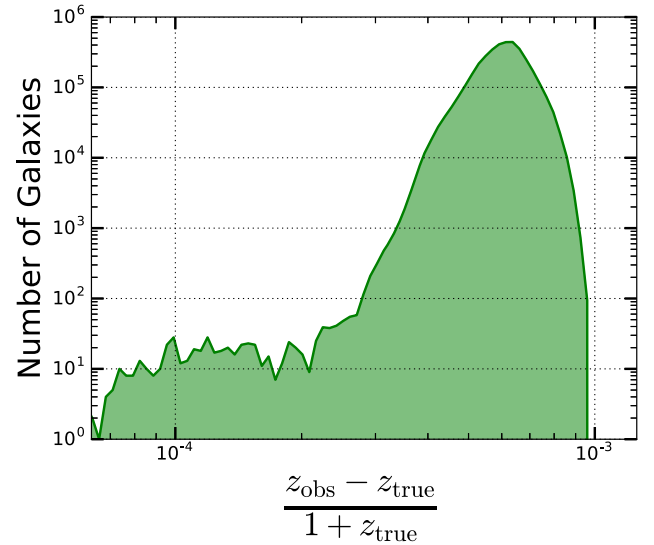


Figure 6. The difference between the observed redshift and the true redshift for a sample of a little over 10 million galaxies. We included only the mean trend in the $[\text{N II}]/\text{H}\alpha$ ratio in the mass–metallicity relationship, but not the scatter. (We expect the mean trend to capture the lowest order correlation between line ratio and environment, but of course excluding the scatter results in an artificially narrow peak.) This does not include the Gaussian photon noise smearing, only the difference due to the line-blending effect. This histogram is generated from one sky section of the parent BUZZARD-V1.1 sample, cut from $z = 0.7$ to $z = 1.3$ (the approximate redshift range of the Z2 bins), and covers approximately 54 square degrees of sky.

and $z \approx 2.3$. We used these empirical relationships to create three linear fits, one for each redshift bin, giving the $[\text{N II}]/\text{H}\alpha$ ratio as a function of log stellar mass. We then used these fits to provide our catalogue with nitrogen line strength ratios, depending on the mass and redshift of each galaxy in our sample. The fits can be seen in Fig. 5.

In our analysis, we make no special allowance for AGNs. We use the ‘All’ fits from Wuyts et al. (2016), so AGN line ratios do pull the fits; however sample sizes in some bins are small, the selection effects may be different from *WFIRST*, and the effect on higher order moments of the $[\text{N II}]/\text{H}\alpha$ distribution (scatter, correlation with other galaxy properties, etc.) are not captured by this procedure, and may be revisited in future work. Furthermore, our analysis is only valid for galaxies on the star-forming main-sequence and not for starbursts, whose $[\text{N II}]/\text{H}\alpha$ ratio could be substantially different due to their high specific SFR with respect to main-sequence galaxies at the same stellar mass and redshift.

Approximately 0.1 per cent of objects in the mock catalogue have very small stellar masses, which when combined with the linear fit from Fig. 5 produce a negative $[\text{N II}]/\text{H}\alpha$ ratio. These ratios are set to zero. This happens for galaxies with $\log_{10}(M/M_{\odot}) < (9.2, 8.9, 9.5)$ for $z \approx (0.9, 1.5, 2.3)$, respectively. We do not expect this simplification to have any effect on our full sample results, due to the relatively small number of galaxies affected,⁶ and because

⁶Most of those low-mass galaxies will not pass the *WFIRST* flux limit ($\sim 10^{-16}$ erg s⁻¹ cm⁻²). Considering a galaxy at the lowest redshift (0.705) and with the limiting flux, with the fiducial cosmology, we estimate the line luminosity to be 2.2×10^{41} erg s⁻¹. The $\text{H}\alpha$ luminosity is lower than that, roughly 1.38×10^{41} erg s⁻¹ if we take the line ratio as 0.6. This provides a conservative estimate for the $\text{H}\alpha$ luminosity, hence the SFR, which is given by $\text{SFR}/[M_{\odot} \text{ yr}^{-1}] \approx 0.57 \times 7.9 \times 10^{-42} L_{\text{H}\alpha}/[\text{erg s}^{-1}] \sim 0.62$ (see

fitting these lower mass galaxies with a more complex model would likely still result in $[\text{N II}]/\text{H}\alpha$ near zero.

4.3 Data binning and redshift distributions

In order to generalize our analysis to many possible future surveys with a range of redshifts and flux thresholds, we binned our simulation catalogue by both redshift and flux, and computed the correlations for all galaxies within each bin. The redshift bins are differential, while the flux bins are integral; their ranges are listed in Table 1. We measure galaxy flux by the *total* emission of the combined $[\text{N II}]$ and $\text{H}\alpha$ lines, which is consistent with the future measurements that will be made by *WFIRST*. In order to decrease computation time, each bin was further divided into six congruent geometric regions on the sky (S1 to S6, displayed in Fig. 3). Correlation functions were independently generated using the counts from each sector.

To ensure that the correlation functions for different groups of measurements, i.e. z_{obs} and z_{true} , can be compared within the exact same redshift limits, it is necessary to use slightly different subsets of galaxies for different calculations – for example one specific galaxy whose true redshift falls in redshift bin 1, may have its observed redshift place it within bin 2. Because of this, the exact galaxy samples vary slightly between true and observed samples.

Each galaxy was given a set of redshifts: a true redshift, an observed redshift, and a shuffled redshift. The true redshift, z_{true} , is simply the original redshift value from the catalogue – that is the redshift that would be observed if the $[\text{N II}]$ and $\text{H}\alpha$ lines could be separately resolved. This redshift still incorporates the peculiar velocities of each galaxy in the redshift value.

We use the observed redshift value, z_{obs} , to include two separate effects. First, we insert the statistical error in the wavelength centroid. Each galaxy’s observed redshift is modified by adding to it a number generated from a Gaussian with mean zero, and standard deviation:

$$\sigma_z = 10^{-3}(1 + z). \quad (2)$$

This error is statistical, and thus not dependent on each individual galaxy’s metallicity. This is the error specified by the *WFIRST* SRD. The real errors will also depend on line flux and galaxy size (σ_z is smaller for galaxies that have brighter lines or smaller angular sizes), however assessment of this is outside the scope of this paper.⁷

The second effect incorporated into z_{obs} is the $[\text{N II}] + \text{H}\alpha$ line-blending effect. As long as the nitrogen line strength is non-zero, this will pull each galaxy’s redshift towards a ‘redder’ value, due to the anisotropy of the nitrogen line pair. This effect is calculated by finding the offset from the $\text{H}\alpha$ line centre, in the rest frame of the galaxy:

$$\delta\lambda = \frac{\Delta_1 F_{6585} - \Delta_2 F_{6550}}{F_{\text{H}\alpha} + F_{6585} + F_{6550}}. \quad (3)$$

equation 10.109 in Mo, van den Bosch & White 2010), where 0.57 is the conversion from Salpeter to Chabrier IMF (Muzzin et al. 2010). Using the SFR–stellar mass relation for the main-sequence galaxies (e.g. Dutton et al. 2010), we find stellar mass to be around $3 \times 10^{10} M_{\odot}$, still above the masses where the ‘negative ratio’ is concerned.

⁷Studying how the statistical variance σ_z^2 depends on galaxy properties will involve both the grism image simulations and (ultimately) *WFIRST* deep field data, which will empirically constrain the precision of the redshift measurement by repeating it many times for the same sample of galaxies.

Here $\delta\lambda$ is the offset of the observed line from the true $\text{H}\alpha$ line, F_{6550} and F_{6585} refer to the strengths of the respective $[\text{N II}]$ lines, $F_{\text{H}\alpha}$ is the strength of the $\text{H}\alpha$ line, Δ_1 is the difference in wavelength between $\text{H}\alpha$ and $[\text{N II}]$ 6585, and Δ_2 is the difference in wavelength between $[\text{N II}]$ 6550 and $\text{H}\alpha$. (Both Δ_1 and Δ_2 are defined to be positive.) We use the vacuum values of 6549.86 Å and 6585.27 Å for the ($^1\text{D}_2^e - ^3\text{P}_1^e$) and ($^1\text{D}_2^e - ^3\text{P}_2^e$) transitions of N II , respectively, and we use a value of $\text{H}\alpha$ at 6564.61 Å. We take the ratio in strengths between the two nitrogen lines, $[\text{N II}]$ 6550/ $[\text{N II}]$ 6585 to be 0.32567 (Storey & Zeippen 2000). Note that equation (3) is valid in the extreme case of a completely unresolved line; the marginally resolved case (relevant for *WFIRST*) can lead to smaller shifts (Faisst et al. 2018), and hence our analysis is conservative.

Once the offset for a specific galaxy is calculated, we can find the observed redshift using

$$z_{\text{obs}} = z_{\text{true}} + \frac{\delta\lambda}{\lambda_{\text{H}\alpha}}(1 + z_{\text{true}}) + \delta_{\text{st}}, \quad (4)$$

where δ_{st} is a realization of the statistical error (see equation 2). Most galaxies show differences between the true redshift and the observed redshift at or below $\delta z = 10^{-3}$ (as shown in Fig. 6), due to the natural $[\text{N II}]/\text{H}\alpha$ ratio empirically found in galaxies. There is a small subset of galaxies, of order 0.1 per cent, with no difference between observed and true redshift, due to the linear metallicity fit. However, the majority of this subset is eliminated during the binning process, as we remove galaxies below a certain flux threshold, and low flux is correlated with the low stellar mass used in metallicity fitting.

After each galaxy has values for z_{true} and z_{obs} , we can then generate the shuffled redshift, z_{shf} . This is done by creating a list of δz values:

$$\delta z \equiv \frac{z_{\text{obs}} - z_{\text{true}}}{1 + z_{\text{true}}}. \quad (5)$$

We then shuffle the list, matching each δz with a true galaxy to create a ‘shuffled’ redshift:

$$z_{\text{shf}} \equiv (1 + z_{\text{true}})\delta z + z_{\text{true}}. \quad (6)$$

This creates a galaxy catalogue where the redshift error distribution is identical to that in the observed catalogue, but where all correlations between δz and galaxy environment are destroyed. In this way, by comparing results from the observed distribution to the shuffled distribution, we can see whether parameter offsets are due to the distribution of redshift errors, or correlations between the redshift error distribution and galaxy environment. This will have an important impact on how we approach mitigation. In Fig. 7, we display the differences in inferred position for a subset of galaxies within the catalogue.

4.4 Correlation functions

Here, we detail our pipeline for calculating the correlation functions within each ZH bin. The pipeline uses the same code as that described in Martens et al. (2018).

Once pair counts were obtained for the true, observed, and shuffled samples within each flux and redshift bin, the redshift distributions of each were used to generate random catalogues. Randomly placed galaxies were created and given redshifts pulled from the distribution of the matching data bin. The random galaxy count is equal to three times the simulated real galaxy count, which was decided on with the intent to minimize error associated

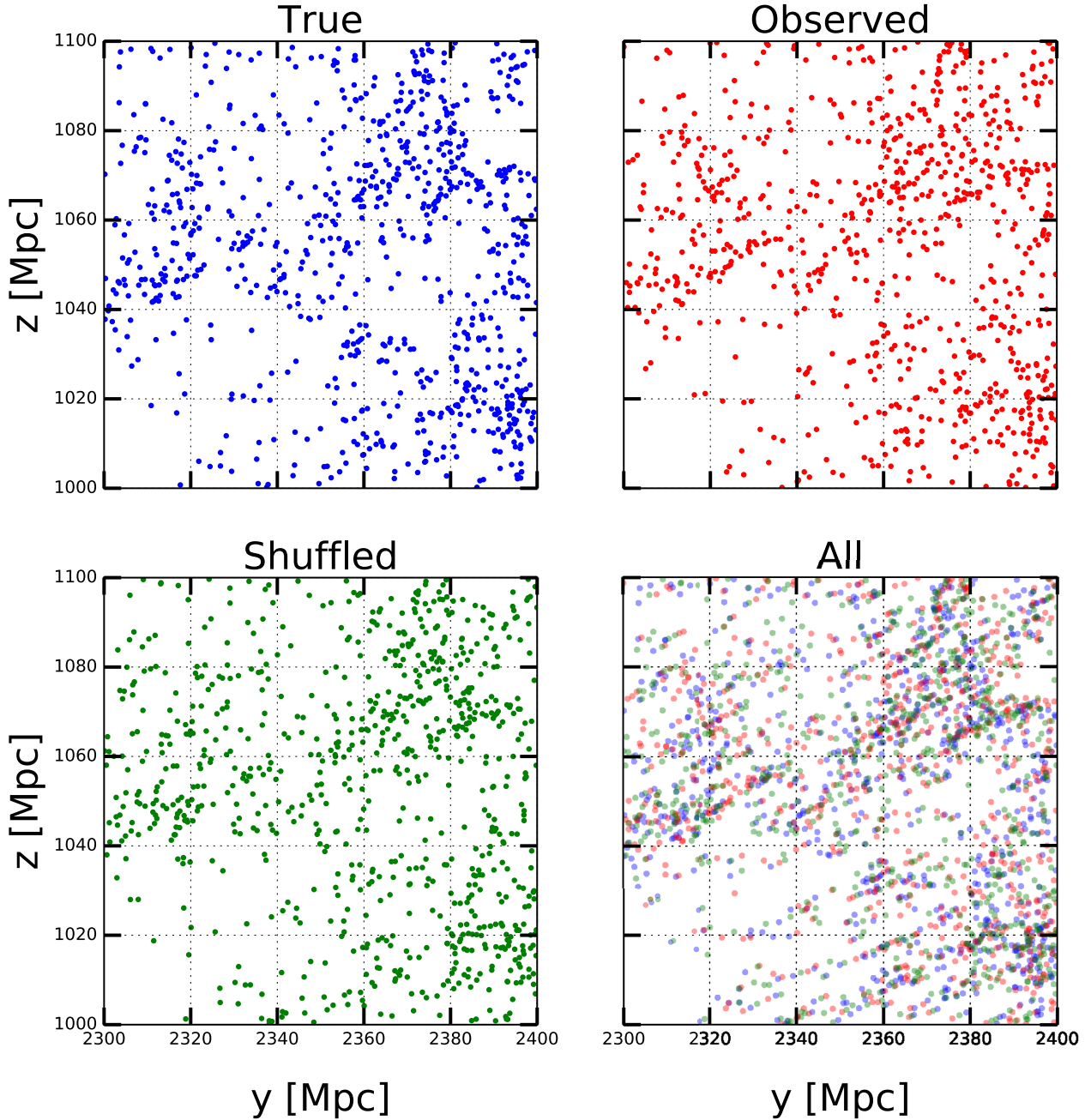


Figure 7. We display the physical positions of the galaxies given the redshifts of each of our true, observed, and shuffled catalogues. These images were generated using Regions 2 and 5 of the Z2H3 catalogues. We have selected all galaxies with a Cartesian x -coordinate (defined by $x = D_c \cos(\text{DEC}) \cos(\text{RA})$, where D_c is the comoving distance) between -50 and 50 Mpc, a z -coordinate (defined by $z = D_c \sin(\text{RA})$) between 1000 and 1100 Mpc, and a y -coordinate (defined by $y = D_c \cos(\text{DEC}) \sin(\text{RA})$) between 2300 and 2400 Mpc. The galaxies in the x -direction have been projected into the y - z plane, and the observer is located at the origin, towards the bottom-left corner of the image (which can be seen by the ‘streaks’ due to the changes in redshift for the combination of the catalogues).

with random galaxy shot noise, but also work within the limits of computation time.⁸ Pair counts were done on the random catalogues in order to construct the correlation function using the Landy–Szalay method (LS: Landy & Szalay 1993; Peebles & Hauser

1974). The correlations are calculated as a function of redshift-space separation s and $\mu = \cos \theta$, where θ is the angle with respect to the line of sight. The LS estimator is

$$\xi(s, \mu) = \frac{\text{DD}(s, \mu) - 2\text{DR}(s, \mu) + \text{RR}(s, \mu)}{\text{RR}(s, \mu)}, \quad (7)$$

where DD refers to the number of pairs of galaxies within a specific distance shell, $s + \Delta s/2$ and $s - \Delta s/2$, and within a specific angular range $\mu + \Delta \mu/2$ and $\mu - \Delta \mu/2$, for the data sample. RR refers to the

⁸For the final *WFIRST* analysis, more resources will be available to devote to random pair counting. The random catalogue shot noise is a fraction n_D/n_R of the shot noise in the data, so to make this negligible, we will need a n_R/n_D much greater than that used in this paper.

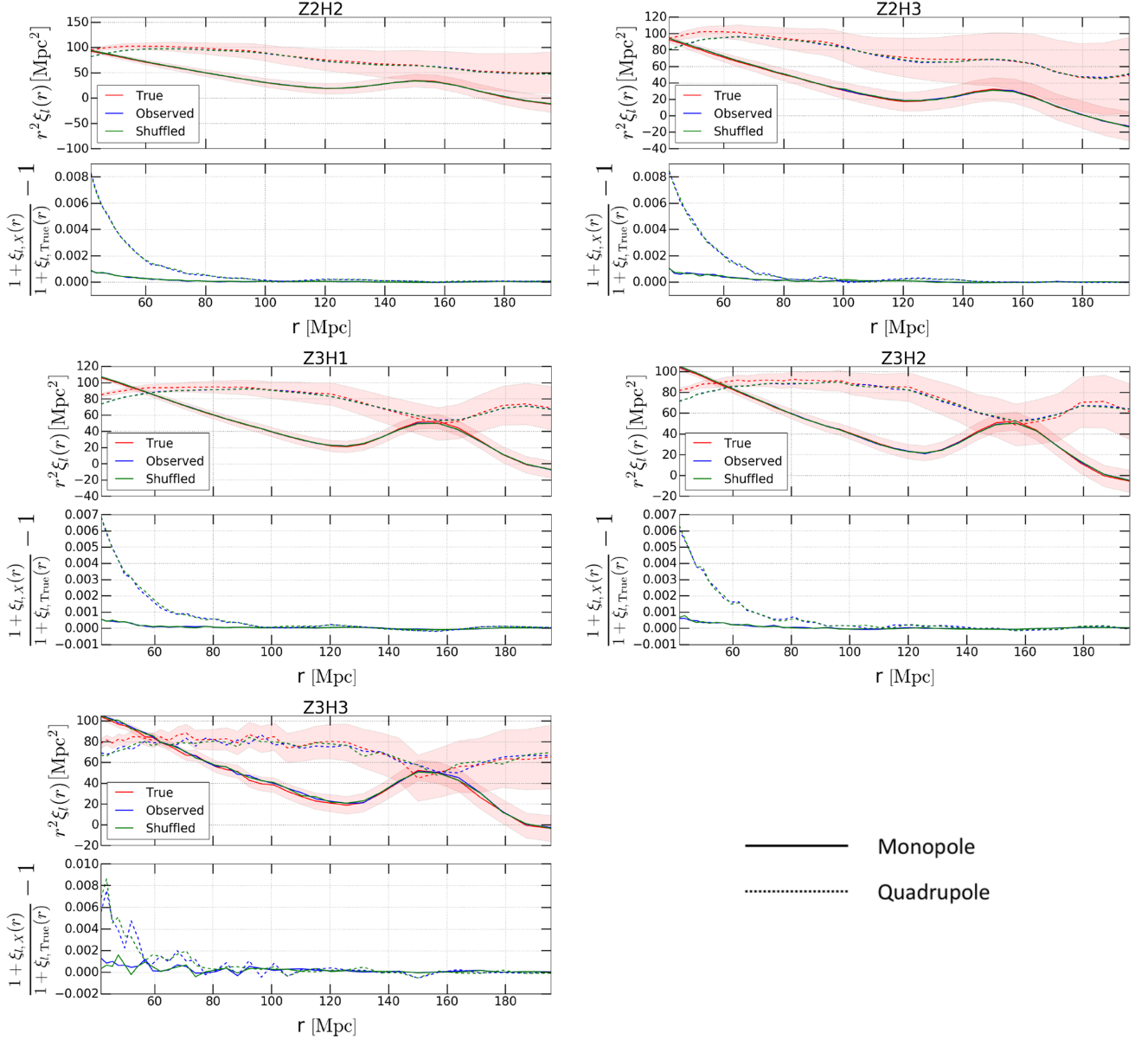


Figure 8. For each ZH bin, we compare the monopole and quadrupole correlation functions for each of the true, observed, and shuffled redshift catalogues. The top panel of each plot displays the correlation function monopoles (solid lines) and quadrupoles (dashed lines). Both quadrupoles and monopoles are multiplied by r^2 , and the quadrupoles are multiplied by -1 . The bottom panel of each plot shows the fractional comparison between the observed and shuffled correlation multipoles, as a function of separation r .

same, but for the random sample. DR refers to a combined catalogue of data and randoms, where we are counting pairs of opposite types only. We use 120 logarithmically spaced radial bins, from $s = 1$ Mpc to $s = 200$ Mpc. Both DD and RR counts are normalized by the total number of galaxies in that bin, specifically:

$$\text{DD} \rightarrow \frac{\text{DD}}{n_D(n_D - 1)} \quad \text{and} \quad \text{RR} \rightarrow \frac{\text{RR}}{n_R(n_R - 1)}, \quad (8)$$

while the DR counts are normalized by:

$$\text{DR} \rightarrow \frac{\text{DR}}{n_D n_R}. \quad (9)$$

Our correlation function code calculates pairs in 20 μ -bins, from -1 to $+1$, with a separation of $\Delta\mu = 0.1$. Although the simulations were generated in μ -binned ‘wedge’ space, we convert them to

multipole space for parameter fitting. The formula for conversion is the same as that used in SDSS BOSS analysis (Ross et al. 2017):

$$\xi_l(r) = \frac{2l+1}{2} \sum_{i=1}^{i_{\max}} \frac{1}{i_{\max}} \xi(r, \mu_i) L_l(\mu_i), \quad (10)$$

where $\mu_i = (i - 1/2)/i_{\max}$, L_l is the Legendre polynomial⁹ of order l , ξ is the μ -binned correlation function, and $i_{\max} = 20/2 = 10$ is the number of μ -bins from 0 to $+1$. In Fig. 8, we plot the resulting monopole and quadrupole correlation functions for the Z2H2, Z2H3, Z3H1, Z3H2, and Z3H3 bins, respectively. In each

⁹To avoid confusion with power spectra, we use L instead of the traditional P for Legendre polynomials.

plot, we show the comparison of the quadrupole and monopole correlation functions between the true, observed, and shuffled redshift catalogues. We also show the fractional comparison from the true catalogue to the observed and shuffled catalogues.

5 PARAMETER FITTING

In this section, we detail the process of fitting our correlation functions from the simulation with smaller scale RSD parameters and large-scale BAO parameters. We perform the RSD and BAO fits completely independently, and in separate scale ranges; although in projects whose primary goal is parameter measurement we would fit both parameter sets together, here we separate them to provide greater sensitivity to the systematic errors we are studying, resulting in the most conservative choice for systematic error budgeting.

First, we use Convolution Lagrangian Perturbation Theory (CLPT, Carlson, Reid & White 2013) and Gaussian Streaming Redshift Space Distortions (GSRSD, Wang, Reid & White 2014) to fit the parameters b_g , the linear galaxy bias, and f_v , the dimensionless linear growth factor, on scales of 42 to 200 Mpc (for varying values of σ_{FOG}^2 , the variance of small-scale dispersion due to FOG, the Finger-of-God effect, in the unit of length^2). This fit focuses on small-scale clustering parameters, and is explained in detail in Section 5.2. Second, we fit BAO parameters on scales of 60 to 200 Mpc, described in Section 5.3. In both cases, we use the correlation function monopole and quadrupole to drive the fitting function.¹⁰

5.1 Covariance matrices

To provide a best-fitting to the observed correlation functions, we assume a Gaussian-distributed likelihood for our vector of measured correlation functions:

$$\mathcal{L}(\mathbf{p}) \propto e^{-\chi^2(\mathbf{p})/2}, \quad (11)$$

where χ^2 is given by:

$$\chi^2(\mathbf{p}) = \sum_{\ell, \ell'} \sum_{i, j} (\xi_{\ell}^i(\mathbf{p}) - \hat{\xi}_{\ell}^i) [\mathbf{C}^{-1}]_{\ell\ell'ij} (\xi_{\ell'}^j(\mathbf{p}) - \hat{\xi}_{\ell'}^j). \quad (12)$$

Here \mathbf{p} is a vector of parameters; ℓ and ℓ' are the moments of the correlation function (here equal to 0 or 2); i, j refer to the separation bins; $\hat{\xi}$ is the measured correlation function; ξ is the model correlation function; and \mathbf{C} is the covariance matrix (Sánchez, Baugh & Angulo 2008; Cohn 2006), which we calculate using the method from Grieb et al. (2016). Grieb et al. (2016) generate a theoretical model for the linear covariance of anisotropic galaxy clustering observations, making use of synthetic catalogues. As input, the calculation of the covariances are based on an input linear galaxy power spectrum dependent on both the wavevector and the angle with the line of sight, $P(k, \mu)$. In order to calculate this, we first calculate the linear matter power spectrum using CLASS (Blas, Lesgourgues & Tram 2011) and then compute the no-wiggle power

¹⁰The scale ranges are chosen for several reasons: (1) the covariance matrix we use is from Grieb et al. (2016), which was tested on scales 30–180 h^{-1} Mpc, i.e. 42–257 Mpc, giving the lower scale cut for the RSD fits; (2) scales 60–200 Mpc is sufficient to cover the BAO peak and the broad-band feature while not being affected by small-scale RSD effects; and (3) cutting the scales at 200 Mpc not only reduces the computation time but also avoids the edge effects due to the finite survey and region areas. We will see in Sections 5.2 and 5.3 that the final results are robust when testing with various scale cuts.

Table 3. Parameters used in generating $P(k, \mu)$ for both the covariance matrix generation and the BAO models used for fitting. These are calculated using formulae from Seo & Eisenstein (2007). For Σ_s , the first value indicates the value used for the true catalogues, while the second value was used for the observed and shuffled catalogues; it was increased due to the extra variance added to simulate photon noise.

Bin	β	$\Sigma_s(\text{Mpc})$	$\Sigma_{\perp}(\text{Mpc})$	$\Sigma_{\parallel}(\text{Mpc})$
Z2H2	0.5762	4.97 (5.46)	7.422	13.709
Z2H3	0.5800	4.99 (5.48)	7.520	13.847
Z3H1	0.4382	4.64 (5.16)	5.679	10.956
Z3H2	0.4424	4.64 (5.16)	5.681	10.959
Z3H3	0.4362	4.64 (5.16)	5.679	10.957

spectrum from the formulae listed in Eisenstein & Hu (1998). This is done at the median redshift of each sample. We next account for RSD effects to the power spectrum using the procedure outlined in Ross et al. (2017):

$$P(k, \mu) = b^2 C^2(k, \mu, \Sigma_s) \left[(P_{\text{nonlin}} - P_{\text{nw}}) e^{-k^2 \sigma_v^2} + P_{\text{nw}} \right], \quad (13)$$

where the no-wiggle power spectrum is also generated with the non-linear power spectrum from the HaloFit model [taking no-wiggle from Eisenstein & Hu (1998) as input] (Smith et al. 2003). We have used

$$\sigma_v^2 = (1 - \mu^2) \Sigma_{\perp}^2 / 2 + \mu^2 \Sigma_{\parallel}^2 / 2 \quad (14)$$

and

$$C(k, \mu, \Sigma_s) = \frac{1 + \mu^2 \beta}{1 + k^2 \mu^2 \Sigma_s^2 / 2}. \quad (15)$$

We define the spreading due to photon noise:

$$\Sigma_{s, \text{phot}} = (300 \text{ km s}^{-1}) \times \frac{1+z}{H(z)}, \quad (16)$$

where we set $\Sigma_s^2 = \Sigma_{s, \text{phot}}^2 + (2.26 \text{ Mpc})^2$. For the true catalogue, $\Sigma_{s, \text{phot}}$ is set to 0. This effectively incorporates the spreading we have added in the shuffled and observed catalogues due to uncertainty in photon noise, and is displayed in Table 3.

In these equations, our values for β , Σ_s , Σ_{\parallel} , and Σ_{\perp} depend on median redshift of the ZH bin we are fitting to, and are listed in Table 3. In order for the matter power spectrum to be used in our covariance matrix calculation, it must be converted to a galaxy power spectrum using a bias appropriate for the tracers which are outlining the dark matter.

To estimate the galaxy bias for each of our samples, we used two separate methods. For the first fitting run only, we used results from the HiZels survey (Cochrane et al. 2017), who perform measurements of the H α emitting galaxies at bins of $z = 0.8, 1.7$, and 2.23, binning further by the mean flux of galaxies in separate bins. We use their estimates of the bias in our covariance matrix (via the power spectrum) to perform the first set of fits. We then record the best-fitting values of the biases from these fits; these values were used for the covariance matrix generation for our second fitting run. This process gave us galaxy biases of approximately [1.47, 1.45, 2.12, 2.10, 2.13] for ZH bins of Z2H2, Z2H3, Z3H1, Z3H2, and Z3H3, respectively. These power spectra are then used to generate covariance matrices in multipole space, for multipoles of 00, 02, 20 (transverse) and 22:

$$C_{l_1, l_2}^{\xi}(s_i, s_j) = \frac{i_{l_1, l_2}^2}{2\pi^2} \int_0^{\infty} k^2 \sigma_{l_1 l_2}^2(k) \bar{J}_{l_1}(ks_i) \bar{J}_{l_2}(ks_j) dk, \quad (17)$$

where the multipole-weighted variance integral is

$$\sigma_{l_1 l_2}^2(k) = \frac{(2l_1 + 1)(2l_2 + 1)}{V_s} \int_{-1}^1 \left[P(k, \mu) + \frac{1}{\bar{n}} \right]^2 L_{l_1}(\mu) L_{l_2}(\mu) d\mu, \quad (18)$$

and the bin-averaged spherical Bessel function is

$$\bar{j}_l(k s_i) = \frac{4\pi}{V_{si}} \int_{s_i - \Delta s/2}^{s_i + \Delta s/2} s^2 j_l(k s) ds. \quad (19)$$

Here $V_{si} = 4\pi[(s_i + \Delta s/2)^3 - (s_i - \Delta s/2)^3]/3$, V_s is the volume of the entire sample, j_l is the spherical Bessel function of the first kind, k is the wavenumber, s is the distance in redshift space, and \bar{n} is the number density of galaxies for the sample in question. In this case, since there is shot noise from both the data and the random catalogues, we make the replacement $1/\bar{n} \rightarrow 1/\bar{n} + 1/\bar{n}_R$ in equation (18); this increases the shot noise by a factor of $\frac{4}{3}$ for $n_R/n_D = 3$.

Our simulations are not volume-limited, and have a galaxy number density, \bar{n} , which is implicitly dependent on redshift, while our theoretical method to calculate the covariance matrices assumes a constant galaxy number density. To account for this, we further divided each of our ZH bins into three sub-bins by redshift. The covariance matrix of each sub-bin was calculated using the volume and number density of galaxies within that specific sub-bin. The covariance matrix of the entire ZH bin was calculated by:

$$C_{l_1 l_2}^\xi(s_i, s_j) = \sum_k w_k^2 C_{l_1 l_2, k}^\xi(s_i, s_j), \quad (20)$$

where k indicates the specific sub-bin, and:

$$w_k = \frac{v_k \bar{n}_k^2}{\sum_{k'} v_{k'} \bar{n}_{k'}^2}, \quad (21)$$

For our Z2 bins, the redshift cut-offs were at 0.9 and 1.1, and for Z3, they were 1.55 and 1.75. The volume of our survey area over the redshift range of the subset k is designated as v_k .

5.2 RSD parameter fitting

The fit on small scales follows the procedure in Martens et al. (2018), focusing on the RSD parameters. We fit on scales of 42 to 200 Mpc, with a factor of four lower spatial resolution than for the BAO fits; it was reduced in order to prevent small-scale fluctuations in the correlation function dominating the best-fitting values.

To calculate the theoretical fit, we use CLPT, validated on small scales by GSRSD. In order to fit a theoretically produced correlation function to our simulations, we use an extension of CLPT (Carlson et al. 2013). CLPT extends perturbation theory beyond linear order to match up to quasi-linear scales of the correlation function; the Gaussian streaming model tailors the fit to behave better on small scales (Wang et al. 2014). GSRSD takes as input σ_{FOG}^2 , f_v , and b_g . We treat σ_{FOG}^2 as a fixed parameter, however we run several fits over different set values of σ_{FOG}^2 , fitting for f_v and b_g with each different set value. This is done to provide a more stringent test to the similarity of the observed and shuffled catalogues, since fixing σ_{FOG}^2 shrinks the uncertainties in the fit for f_v ; it should be noted that the real data collected by WFIRST will be simultaneously fit for f_v , b_g , and σ_{FOG}^2 , although the scales over which these will be fit could differ from those presented here.

The observed and shuffled catalogues have a different set value of σ_{FOG}^2 than the true catalogues, due to the spreading introduced in equation (2). This additional spreading was calculated separately for

each ZH bin, and was found to be 24.222, 24.324, 21.097, 21.098, and 21.080 $h^{-2} \text{Mpc}^2$ for ZH bins of Z2H2, Z2H3, Z3H1, Z3H2, and Z3H3, respectively. We run our fits for σ_{FOG}^2 values of 5, 20, and 35 ($h^{-1} \text{Mpc}$)² for the true catalogues, which is added to the additional spreading found for the observed and shuffled catalogues.

In order to produce these outputs, GSRSD also takes as input the galaxy bias, and f_v , which are our primary fitting parameters. The code outputs the redshift-space correlation function in terms of moments, $\xi_{0, 2, 4}(r)$, which is directly comparable to our simulated correlation functions.

We have repeated the RSD parameters for alternative scale ranges (e.g. 60–180 Mpc) and found that the true/observed/shuffled parameter shifts change by < 0.1 per cent.

5.3 BAO parameters

WFIRST will perform accurate BAO measurements up to $z \sim 1.9$ using H α , and to higher redshifts using [O III] and [O II] emitters, pinning down the expansion rate of the Universe, $H(z)$ and the angular diameter distance, $D_A(z)$. However, the line-blending effect studied in this paper will potentially bias the results. To quantify this bias and track down how much arises from the redshift error PDF versus its correlation with large-scale structure, we will compare the BAO parameters measured from our true, observed, and shuffled catalogues. In this subsection, we first introduce the BAO model, and then discuss about our fitting process. They do *not* include reconstruction, which we leave to future analysis.

The BAO fits are performed over scales from 60 to 200 Mpc, and are intended to calculate the expansion rate of the Universe at a specific redshift, $H(z)$, and the angular diameter distance to that redshift, $D_A(z)$. We follow the standard BAO convention in defining

$$\alpha_{\parallel} = \frac{(H(z)r_d)^{\text{fid}}}{H(z)r_d} \quad \text{and} \quad \alpha_{\perp} = \frac{D_A(z)r_d^{\text{fid}}}{D_A(z)r_d}, \quad (22)$$

where here, a superscript of ‘fid’ indicates a fiducial value, and r_d is the sound horizon at the kinetic decoupling epoch. Given $P(k, \mu)$, we calculate the multipole moments

$$P_l(k) = \frac{2l+1}{2} \int_{-1}^1 P(k, \mu) L_l(\mu) d\mu, \quad (23)$$

and then transform them into real space as

$$\xi_l(s) = \frac{i^l}{2\pi^2} \int \frac{dk}{k} k^3 P_l(k) j_l(k s), \quad (24)$$

so that the correlation function is expressed as

$$\xi(s, \mu) = \sum_l \xi_l(s) L_l(\mu), \quad (25)$$

where we only sum to $l = 4$.

We fit to the same BAO model in Ross et al. (2017),

$$\xi_{0, \text{mod}}(s) = B_0 \xi_{\mu 0}(s, \alpha_{\perp}, \alpha_{\parallel}) + \frac{a_{01}}{s^2} + \frac{a_{02}}{s} + a_{03} \quad \text{and} \quad (26)$$

$$\begin{aligned} \xi_{2, \text{mod}}(s) = & \frac{5}{2} [B_2 \xi_{\mu 2}(s, \alpha_{\perp}, \alpha_{\parallel}) - B_0 \xi_{\mu 0}(s, \alpha_{\perp}, \alpha_{\parallel})] \\ & + \frac{a_{21}}{s^2} + \frac{a_{22}}{s} + a_{23}, \end{aligned} \quad (27)$$

where $\xi_{\mu 0}$, $\xi_{\mu 2}$ are μ -averaged $\xi(s, \mu)$, defined by

$$\xi_{\mu 0}(s, \alpha_{\perp}, \alpha_{\parallel}) = \int_0^1 d\mu \xi(s', \mu') \quad (28)$$

and

$$\xi_{\mu 2}(s, \alpha_{\perp}, \alpha_{\parallel}) = \int_0^1 d\mu \, 3\mu^2 \xi(s', \mu'), \quad (29)$$

with $\mu' = \mu\alpha_{\parallel}/\sqrt{\mu^2\alpha_{\parallel}^2 + (1 - \mu^2)\alpha_{\perp}^2}$, and $s' = s\sqrt{\mu^2\alpha_{\parallel}^2 + (1 - \mu^2)\alpha_{\perp}^2}$. Parameters B_i and a_{ij} set the size of the BAO and the broad-band feature.

We fit the model by minimizing the χ^2 in equation (12). Since the model depends linearly on the parameters B_i and a_{ij} , the χ^2 is a quadratic function of them. For each given pair of $(\alpha_{\perp}, \alpha_{\parallel})$, one can calculate the other eight parameters where the χ^2 takes a minimum. Let $\mathbf{p}^{(8)}$ be the vector of the eight parameters, and p_i be its i -th component; then the χ^2 can be written as

$$\chi^2(\alpha^{\perp}, \alpha^{\parallel}, \mathbf{p}^{(8)}) = \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d} - 2 \sum_{i=1}^8 p_i J_i + \sum_{i,j=1}^8 p_i p_j K_{ij}. \quad (30)$$

Here \mathbf{d} is the data vector, and we define $J_i = \mathbf{m}_i^T \mathbf{C}^{-1} \mathbf{d}$ and $K_{ij} = \mathbf{m}_i^T \mathbf{C}^{-1} \mathbf{m}_j$, where \mathbf{m}_i is the model vector if the parameter p_i is set to be 1 and all other seven components of $\mathbf{p}^{(8)}$ are set to be 0. The minimum of χ^2 is obtained if $p_i^{(8)} = [\mathbf{K}^{-1}]_{ij} J_j$. We first use equation (24) to generate the theoretical $\xi_{0,2,4}(s)$ with 5000 s -values logarithmically spaced in $[1, 240]$ Mpc for each ZH bin. Several techniques have been applied to improve the accuracy of the integration. First, we use 50 000 k -values logarithmically sampled from 10^{-4} to 100 Mpc^{-1} . Secondly, a window function that has continuous first and second derivatives¹¹ is applied to the high- k end of the k -dependent function $k^3 P_l(k)$, to remove the high-frequency ringing appearing in $\xi_l(s)$. Finally, we multiply $j_l(ks)$ by a factor of $\sin(ks\Delta \ln k/2)/(ks\Delta \ln k/2)$ to effectively average out the contribution from the rapidly oscillating spherical Bessel functions at large k . During each iteration of the minimization, $\xi_l(s')$ is calculated from the precalculated theoretical $\xi_l(s)$ using the cubic interpolation method. The minimization uses the `scipy.optimize.minimize` routine (Jones et al. 2001) with the Nelder–Mead method, and the initial guess of $(\alpha_{\perp}, \alpha_{\parallel})$ is always (1,1).

We have tried adjusting the BAO fitting range, varying the lower cut-off from the default (60 Mpc) to 50, 70, or 80 Mpc; the BAO scale parameters change by < 0.1 per cent in all cases.

5.4 Fitting results

Here, we show the resulting parameters that were fit to each correlation function, and discuss trends in the results. For each ZH bin, we have six separate regions, which serve as six realizations of simulation data. We average the correlation functions from all sectors:

$$\xi_{\text{avg}} = \frac{\xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5 + \xi_6}{6}. \quad (31)$$

We then fit parameters to ξ_{avg} , which are denoted as \bar{p} for a given parameter p . These fits provide the listed parameters in Tables 4 and 5, as well as the listed χ^2 per degree of freedom.

We find the error bars for each parameter using the jack-knife method, where we construct six separate combinations of the correlation functions, with each combination containing all regions

except for one:

$$\bar{\xi}_i = \frac{1}{5} \sum_{j=1, j \neq i}^6 \xi_j. \quad (32)$$

We then separately fit the correlation functions $\bar{\xi}_i$ for the BAO and RSD parameters. We show error bars which are derived from the variance of the set of parameters of our combinations:

$$\sigma_{\bar{p}} = \sqrt{\frac{5}{6} \sum_{i=1}^6 (\bar{p}_i - \bar{p})^2}. \quad (33)$$

The jack-knife method, while providing an estimate of the errors in our parameters due to deviations in the correlation functions, is limited by our small sample set of only six realizations. Table 4 displays the f_v and b_g best-fitting parameters for each ZH bin, and each of the true, observed, and shuffled catalogues. There are three columns in the table, each indicating the best fits in cases of varying values for the Finger-of-God spreading, σ_{FOG}^2 . We list the true catalogue value of σ_{FOG}^2 for the true catalogue in Table 4. These values were chosen to represent a wide range of values for σ_{FOG}^2 for our samples, and to show how their choice affects the differences between the observed and shuffled catalogues. The error bars here are calculated using equation (33). Table 4 also shows, for each ZH bin, the percent error between the fitted parameters between the observed and shuffled catalogues. This is calculated by:

$$\Delta p \equiv \frac{p_{\text{obs}} - p_{\text{shf}}}{0.5(p_{\text{obs}} + p_{\text{shf}})} \times 100 \text{ per cent}, \quad (34)$$

where p is either b_g, f_v, α_{\perp} , or α_{\parallel} .

In Fig. 9, we plot the best-fitting parameters for a visual comparison of the different RSD results, we show the full RSD best-fitting parameters in Table 4, and the best-fitting BAO parameters are displayed in Table 5.

Systematic errors in the mean parameter fits could arise from a few assumptions made in the modelling. Mean redshifts were calculated for each ZH bin, and were used to determine the parameters Σ_{\perp} , Σ_{\parallel} , Σ_s , and generate the matter power spectra. Since our samples are not volume-limited, the redshift dependence of the galaxy number densities in each bin may introduce errors in the modelling.

Some of the large $\chi^2/\text{d.o.f.}$ values result from the combination of the bin-to-bin scatter in $\xi_l(s)$ and the strong correlations between neighbouring radial bins (i.e. the off-diagonal terms of the covariance matrix). We suspect that the analytic covariance matrix may not be adequate for describing these rapidly oscillating modes in $\xi_l(s)$. Although we attempt to generalize it to a non-volume-limited sample, we still must estimate a galaxy bias for the input power spectrum, and the rapidly oscillating modes correspond to high k where non-linear biasing may be important. The covariance matrix uses a Gaussian approximation, neglecting the galaxy trispectrum and supersample variance effects, which may also break down for these modes. In any case, these highly oscillating modes are orthogonal to the broad-band modes that dominate f_v , and changes in our treatment of the covariance matrix in the preparation of this paper had a small impact on changes in the parameter shifts.

6 MEASURING AND CORRECTING ERROR TERMS

As discussed in Section 2, *WFIRST* has a requirement to meet, and preferably exceed, the observational systematic error requirements

¹¹See equation (C.1) in McEwen et al. (2016). Here we have $k_{\text{max}} = 100 \text{ Mpc}^{-1}$ and the window applies from 80 to 100 Mpc^{-1} .

Table 4. Here we display the best-fitting parameters for the growth of structure parameter f_v and the galaxy bias b_g for each ZH bin and redshift catalogue. We calculate per cent error (equation 34) to indicate the agreement between the shuffled and observed parameters. The uncertainties provided by the jack-knife sampling have a fractional error of $\sqrt{2/5}$, due to the sample size of six regions.

	b_g	$\sigma_{\text{FOG}}^2 = 5 (h^{-1} \text{Mpc})^2$ f_v	$\chi^2/\text{d.o.f}$	b_g	$\sigma_{\text{FOG}}^2 = 20 (h^{-1} \text{Mpc})^2$ f_v	$\chi^2/\text{d.o.f}$	b_g	$\sigma_{\text{FOG}}^2 = 35 (h^{-1} \text{Mpc})^2$ f_v	$\chi^2/\text{d.o.f}$
Z2H2									
True	1.4817 \pm 0.0139	0.7080 \pm 0.0148	1.219	1.4658 \pm 0.0139	0.7593 \pm 0.0147	1.285	1.4487 \pm 0.0137	0.8129 \pm 0.0149	1.819
Observed	1.4938 \pm 0.0120	0.6878 \pm 0.0126	1.674	1.4729 \pm 0.0119	0.7489 \pm 0.0127	1.359	1.4497 \pm 0.0119	0.8135 \pm 0.0128	1.813
Shuffled	1.4934 \pm 0.0120	0.6898 \pm 0.0113	1.618	1.4724 \pm 0.0118	0.7504 \pm 0.0115	1.336	1.4493 \pm 0.0119	0.8154 \pm 0.0115	1.828
Per cent error	0.03 \pm 0.10 per cent	-0.28 \pm 0.32 per cent	–	0.04 \pm 0.12 per cent	-0.20 \pm 0.30 per cent	–	0.03 \pm 0.11 per cent	-0.22 \pm 0.25 per cent	–
Z2H3									
True	1.4585 \pm 0.0139	0.6891 \pm 0.0200	1.512	1.4429 \pm 0.0138	0.7380 \pm 0.0199	1.574	1.4262 \pm 0.0139	0.7892 \pm 0.0201	1.971
Observed	1.4643 \pm 0.0107	0.6734 \pm 0.0161	1.464	1.4444 \pm 0.0107	0.7298 \pm 0.0161	1.376	1.4227 \pm 0.0107	0.7898 \pm 0.0164	1.803
Shuffled	1.4676 \pm 0.0127	0.6707 \pm 0.0178	1.617	1.4479 \pm 0.0125	0.7271 \pm 0.0179	1.478	1.4262 \pm 0.0125	0.7871 \pm 0.0180	1.847
Per cent error	-0.22 \pm 0.37 per cent	0.40 \pm 0.51 per cent	–	-0.25 \pm 0.39 per cent	0.36 \pm 0.49 per cent	–	-0.24 \pm 0.38 per cent	0.34 \pm 0.41 per cent	–
Z3H1									
True	2.1512 \pm 0.0072	0.7631 \pm 0.0065	2.373	2.1294 \pm 0.0074	0.8287 \pm 0.0066	2.112	2.1056 \pm 0.0073	0.8974 \pm 0.0066	2.949
Observed	2.1531 \pm 0.0054	0.7561 \pm 0.0076	2.525	2.1256 \pm 0.0052	0.8305 \pm 0.0078	1.956	2.0956 \pm 0.0054	0.9090 \pm 0.0079	2.992
Shuffled	2.1539 \pm 0.0060	0.7574 \pm 0.0064	2.362	2.1268 \pm 0.0059	0.8317 \pm 0.0064	1.952	2.0967 \pm 0.0058	0.9104 \pm 0.0065	3.156
Per cent error	-0.04 \pm 0.09 per cent	-0.18 \pm 0.41 per cent	–	-0.06 \pm 0.09 per cent	-0.14 \pm 0.41 per cent	–	-0.05 \pm 0.07 per cent	-0.15 \pm 0.36 per cent	–
Z3H2									
True	2.1307 \pm 0.0066	0.7476 \pm 0.0070	2.149	2.1095 \pm 0.0066	0.8093 \pm 0.0073	1.939	2.0867 \pm 0.0063	0.8740 \pm 0.0073	2.593
Observed	2.1335 \pm 0.0045	0.7483 \pm 0.0087	2.095	2.1079 \pm 0.0045	0.8170 \pm 0.0088	1.939	2.0798 \pm 0.0043	0.8894 \pm 0.0092	2.980
Shuffled	2.1350 \pm 0.0039	0.7467 \pm 0.0097	2.211	2.1095 \pm 0.0039	0.8156 \pm 0.0100	2.069	2.0812 \pm 0.0038	0.8881 \pm 0.0100	3.121
Per cent error	-0.07 \pm 0.10 per cent	0.21 \pm 0.34 per cent	–	-0.08 \pm 0.11 per cent	0.17 \pm 0.33 per cent	–	-0.07 \pm 0.10 per cent	0.15 \pm 0.29 per cent	–
Z3H3									
True	2.1535 \pm 0.0121	0.7217 \pm 0.0112	0.946	2.1332 \pm 0.0124	0.7737 \pm 0.0112	1.260	2.1112 \pm 0.0123	0.8276 \pm 0.0114	1.945
Observed	2.1650 \pm 0.0105	0.7152 \pm 0.0116	1.425	2.1420 \pm 0.0104	0.7710 \pm 0.0117	1.762	2.1171 \pm 0.0103	0.8290 \pm 0.0120	2.543
Shuffled	2.1616 \pm 0.0126	0.7079 \pm 0.0109	1.365	2.1390 \pm 0.0126	0.7632 \pm 0.0111	1.522	2.1143 \pm 0.0127	0.8214 \pm 0.0114	2.118
Per cent error	0.15 \pm 0.30 per cent	1.03 \pm 1.21 per cent	–	0.14 \pm 0.30 per cent	1.01 \pm 1.11 per cent	–	0.13 \pm 0.31 per cent	0.92 \pm 1.09 per cent	–

Table 5. The best-fitting parameters for the BAO shift parameters α_{\parallel} and α_{\perp} , and their χ^2 values. We calculate per cent error (equation 34) to indicate the agreement between the shuffled and observed parameters. The uncertainties provided by the jack-knife sampling have a fractional error of $\sqrt{2/5}$, due to the sample size of six regions.

BAO Parameters			
	α_{\parallel}	α_{\perp}	$\chi^2/\text{d.o.f}$
Z2H2			
True	0.9858 ± 0.0111	0.9800 ± 0.0066	0.484
Observed	0.9920 ± 0.0092	0.9780 ± 0.0062	0.678
Shuffled	0.9895 ± 0.0098	0.9792 ± 0.0054	0.807
Per cent error	0.26 ± 0.17 per cent	-0.12 ± 0.11 per cent	–
Z2H3			
True	1.0045 ± 0.0166	0.9739 ± 0.0088	0.486
Observed	0.9977 ± 0.0154	0.9708 ± 0.0076	0.826
Shuffled	0.9986 ± 0.0153	0.9726 ± 0.0078	0.871
Per cent error	-0.08 ± 0.24 per cent	-0.19 ± 0.11 per cent	–
Z3H1			
True	0.9614 ± 0.0048	0.9763 ± 0.0038	1.537
Observed	0.9598 ± 0.0044	0.9799 ± 0.0039	1.362
Shuffled	0.9586 ± 0.0051	0.9793 ± 0.0041	1.241
Per cent error	0.13 ± 0.19 per cent	0.07 ± 0.05 per cent	–
Z3H2			
True	0.9592 ± 0.0025	0.9753 ± 0.0041	1.668
Observed	0.9668 ± 0.0042	0.9690 ± 0.0035	1.427
Shuffled	0.9638 ± 0.0025	0.9699 ± 0.0037	1.468
Per cent error	0.31 ± 0.23 per cent	-0.10 ± 0.10 per cent	–
Z3H3			
True	0.9625 ± 0.0110	0.9757 ± 0.0033	1.286
Observed	0.9488 ± 0.0088	0.9702 ± 0.0064	1.109
Shuffled	0.9504 ± 0.0080	0.9721 ± 0.0060	1.189
Per cent error	-0.17 ± 0.55 per cent	-0.19 ± 0.15 per cent	–

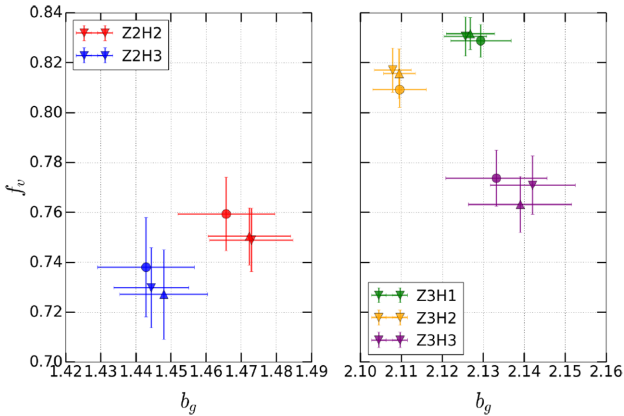


Figure 9. For each ZH bin and observed, true, and shuffled catalogue, we plot the fitted values for f_v and b_g . The circles indicate the true catalogue, V's indicate the observed catalogue, and triangles indicate the shuffled catalogue. For this fitting calculation, σ_{FOG}^2 was set to $20[h^{-1}\text{Mpc}]^2$ for the true catalogue.

in the SRD for measured parameters; specifically, the parameters measured in this work are the BAO shift parameters α_{\parallel} and α_{\perp} , and the growth of structure parameter f_v . The error limits are 0.41 per cent for α_{\perp} , 0.74 per cent for α_{\parallel} , and 0.74 per cent for f_v (there is no top-level requirement on the systematic errors to

the galaxy bias b_g , as we will marginalize over it in estimating cosmological parameters). In this section, we discuss the breakdown of line-blending errors within the structure of our catalogues, compare our results to the error budget of the SRD, and briefly introduce potential mitigation techniques that could limit the negative effects of line blending on the observed redshift. Note that in our discussions of ‘per cent of error budget’ used, we will use root-sum-square (RSS) error budgeting, so that an effect whose amplitude is 50 per cent of the maximum allowed error is considered to use 25 per cent of the total budget.

In this paper, we have presented the fit parameters for three separate catalogues of redshift: true, observed, and shuffled. Their derivation is described in detail in Section 4.3. The purpose of these catalogues is to separate the redshift errors introduced by line blending into two distinct sources. The difference between the parameters of the observed and true catalogues (which we wish to eventually mitigate) can be written as:

$$p_{\text{obs}} - p_{\text{true}} = (p_{\text{obs}} - p_{\text{shuffled}}) + (p_{\text{shuffled}} - p_{\text{true}}). \quad (35)$$

The first part of equation (35) is comprised of the difference between the parameters of the observed and shuffled catalogues. By construction, these two catalogues have identical *distributions* of redshift errors; however, in the shuffled catalogue, the line-blending redshift difference has been uncoupled from the specific galaxy that generated it, destroying the correlation between [N II]/H α and galaxy environment. Therefore, this term describes the effect of this

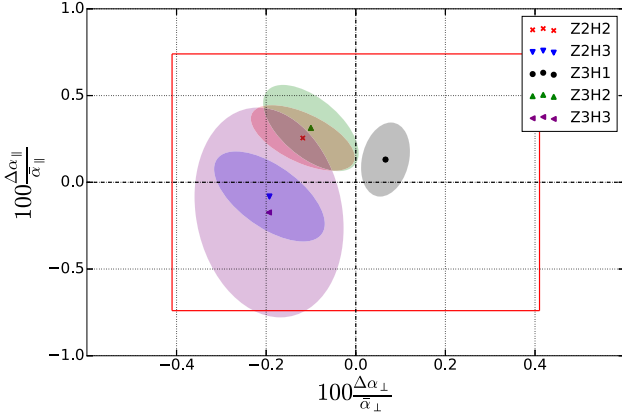


Figure 10. For each ZH bin, we show the percent difference of α_{\perp} and α_{\parallel} relative to the systematic error budget (red box) of *WFIRST* for each parameter. The contours show the spread of fits for all six jack-knife combinations for that specific ZH bin, as referenced in equation (33), while the central values are calculated from the fits of the average of all regions.

correlation. It is this term that – although small – is not amenable to mitigation by measuring the one-point PDF of galaxy properties.

In contrast, the second part of equation (35) depends only on the one-point properties of galaxies, such as their metallicity distribution. Because there is *no* large-scale structure information encoded in this term, it is more straightforward to mitigate via detailed observations of a small number of galaxies. We will analyse the error encompassed by these terms separately, in order to discuss the necessary mitigation techniques.

6.1 Effect of correlations between [N II]/H α and large-scale environment

We first analyse the difference between the shuffled and observed catalogues. Although we have fit a variety of redshift and flux bins, we will confine our numerical analysis here to the Z2H2 and Z3H2 bins, since the limiting flux of $1.3 \times 10^{-16} \text{ erg cm}^{-2} \text{ s}^{-1}$ is closest to the planned *WFIRST* flux limit.¹² For the BAO parameters α_{\perp} and α_{\parallel} , the best-fitting values are listed in Table 5, along with the percentage difference between the observed and shuffled best-fitting values. These differences are also shown graphically in Fig. 10. The magnitude of the differences in α_{\parallel} fall generally in the range 0.1–0.3 per cent, while for α_{\perp} they fall in the range 0.1–0.2 per cent. The errors on these differences are provided by the jack-knife method for all six regions we fit for. It is important to note that the jack-knife errors are based on a small sample size of six regions (for which the expected fluctuations in error bars are $\pm 1/\sqrt{2(6-1)} \approx 32$ per cent); this could explain why in some cases we have a larger error bar for a ZH bin with a smaller sample size.

Specifically, we find that $\Delta\alpha_{\parallel} = 0.31 \pm 0.23$ per cent (0.26 ± 0.17 per cent) for Z3H2 (Z2H2), and $\Delta\alpha_{\perp} = -0.10 \pm 0.10$ per cent (-0.12 ± 0.11 per cent) for Z3H2 (Z2H2). Using the largest value of each, this corresponds to $\Delta\alpha_{\parallel}$ using 18 per cent of the systematic error budget, while $\Delta\alpha_{\perp}$ uses 9 per cent. This is a small percentage of the systematic error budget, especially considering that it is for the unmitigated result.

¹²This flux limit varies somewhat as a function of wavelength and ecliptic latitude, and is subject to change during future optimization.

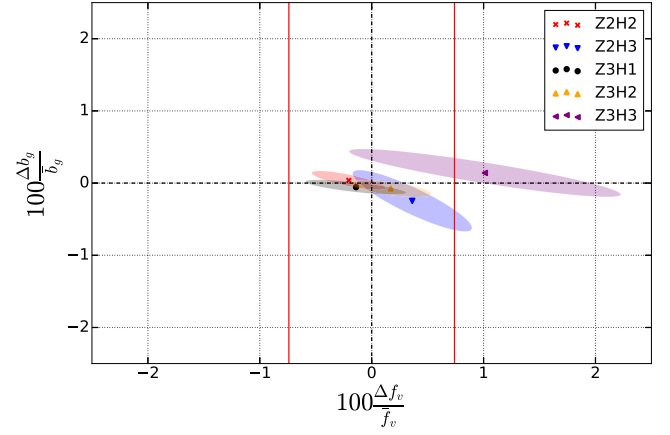


Figure 11. For each ZH bin, we show the percent difference of f_v and b_g relative to the systematic error budget of *WFIRST* for f_v , represented by the solid red lines. For this fitting calculation, σ_{FOG}^2 was set to $20[h^{-1}\text{Mpc}]^2$ for the true catalogue. The contours show the spread of fits for all six jack-knife combinations for that specific ZH bin, as referenced in equation (33), while the central values are calculated from the fits of the average of all regions.

We can also compare these errors to a previous estimate of the line-blending effect: Faisst et al. (2018) looked at the overall redshift errors due to line blending, and propagated it the calculation of the BAO parameters; they predicted an upper estimate of these errors to be in the range of 0.5–1.6 per cent for the α_{\parallel} and α_{\perp} parameters. Our clustering analysis has shown that this estimate is indeed an upper limit, as the values we have found due to clustering are significantly lower.

The error from this portion of equation (35) can be potentially reduced if the [N II]/H α ratio can be predicted from other available data. For example, the broad-band LSST + *WFIRST* photometry (which extends into the rest-frame optical) allows estimates of the stellar mass M_* , and observations of a small fraction of the *WFIRST* sources with high-resolution ground-based NIR spectrographs would enable the correlation between [N II]/H α ratio to be determined as a function of redshift z and inferred stellar mass M_* . After applying this correction, the remaining systematic error from line blending would be associated only with the residuals from the [N II]/H α versus (z , M_*) fit and their correlation with large-scale environment. Future work will be required to determine how much the systematic errors can be mitigated by this method.

We display the best-fitting parameters for the RSD fits in Table 4, with the percentage difference between the observed and shuffled best-fitting values also displayed in Fig. 11. We have fit f_v given several different values of σ_{FOG}^2 , and found consistency in the percentage errors in each case; here we use the values from the fit for the $\sigma_{\text{FOG}}^2 = 20 h^{-2}\text{Mpc}^2$ bin. The differences in f_v fall in the range 0.14–0.36 per cent, with the exception of the Z3H3 bin, which has an error of 1.01 ± 1.11 per cent (the error bar is large due to the high flux limit and corresponding small sample size). For purposes of comparing to the SRD, we find that the error for f_v is 0.17 ± 0.33 per cent (-0.20 ± 0.30 per cent) for Z3H2 (Z2H2). For Z2H2, the larger difference, this corresponds to f_v using 7 per cent of the systematic error budget, which again is a small enough offset that mitigations may not be required. If necessary, the aforementioned mitigations based on measurement of the [N II]/H α versus (z , M_*) relation could reduce it even further.

6.2 Effects of the one-point redshift error PDF

Next, we evaluate the differences between the true and shuffled catalogues, which constitutes the second term in equation (35). We find that the difference for α_{\parallel} is approximately -0.48 per cent (-0.37 per cent) for the Z3H2 (Z2H2) bin, which constitutes 42 per cent of the systematic error (for the largest case). This error is similar for α_{\perp} at 0.56 per cent (0.08 per cent) for Z3H2 (Z2H2), but the lower error budget means it constitutes 187 per cent of the total allotment (again for the largest case). In order to move these errors below 25 per cent of the SRD limits, this would require a reduction in the errors by factors of approximately 1.3 and 2.8 for α_{\parallel} and α_{\perp} , respectively. For f_v , the difference is -0.78 per cent (1.18 per cent) for Z3H2 (Z2H2), resulting in a maximum value of 254 per cent of the total allotment. This requires a reduction in the errors by a factor of about 3.2 to get within this 25 per cent limit.

This could be achieved with a spectroscopic reobservation of some subset of galaxies already observed by *WFIRST* with a high-resolution ground-based spectrograph that completely resolves $H\alpha$ from the [N II] doublet. Only a tiny fraction of *WFIRST* source could be followed up this way, but it would provide a clean measurement of the redshift error PDF. In this way, we gain knowledge of each reobserved object's specific z_{true} , and can construct a probability distribution $P_{\text{est}}(\delta_z|z_{\text{obs}})$ based on these reobserved objects. The more objects we are able to reobserve, the more our subsample approaches the full sample, and the more our probability distribution $P_{\text{est}}(\delta_z|z_{\text{obs}})$ approaches the true distribution $P(\delta_z|z_{\text{obs}})$.

In the case of f_v , we predict how many galaxies would need to be reobserved in order to make this reduction. From Table 4 it can be seen that (for both Z2H2 and Z3H2):

$$\frac{\partial \ln f_v}{\partial \sigma_{\text{FOG}}^2} \approx 0.005 (h^{-1} \text{Mpc})^{-2}. \quad (36)$$

This quantifies the dependence of f_v on the Finger-of-God length; if we place limits on f_v , we need to know σ_{FOG}^2 to some certainty as well, since they are correlated in their fit values. Therefore, to fit within 25 per cent of the error budget, we need σ_{FOG}^2 to be known to:

$$\begin{aligned} \sigma_{\text{FOG, err}} &= \left(\frac{\partial \ln f_v}{\partial \sigma_{\text{FOG}}^2} \right)^{-1} \times f_{v, \text{err}} \sqrt{0.25} \\ &= \frac{0.0074 \sqrt{0.25}}{0.005} = 0.75 h^{-2} \text{Mpc}^2. \end{aligned} \quad (37)$$

Given the variance of the redshift offset δz , which corresponds to $24.2 h^{-2} \text{Mpc}^2$ for Z2H2 and $21.1 h^{-2} \text{Mpc}^2$ for Z3H2, then we need to know σ_{FOG}^2 to about $0.75/24.2 = 3.1$ per cent for Z2H2 and $0.75/21.1 = 3.6$ per cent for Z3H2. If assuming Gaussian errors, this would require an observation approximately in the range of $2/0.031^2 \approx 2100$ galaxies. In practice, the number of reobservations would be larger for a realistic non-Gaussian error distribution, and a few redshift bins would be necessary to track the redshift dependence of the error PDF.

Additional work will be necessary to determine the optimal strategy for measuring the redshift error PDF. Following up thousands of emission-line galaxies with ground-based NIR spectroscopy at the $\sim 10^{-16} \text{ erg cm}^{-2} \text{ s}^{-1}$ depth is certainly possible, especially if the targets can be pre-selected from *WFIRST* to have lines that will not collide with atmospheric OH features. However, since most of the redshift error is statistical error due to photon noise, it may be more efficient to use repeat observations in the *WFIRST* deep fields to measure the purely statistical scatter, and then high-resolution

ground-based NIR spectra to constrain the specific contribution from line blending. These possibilities should be explored in future work.

7 DISCUSSION

In this paper, we have examined the effects of the grism resolution proposed for *WFIRST* on the observed redshifts of galaxies, and the resultant changes in the fitted cosmological parameters α_{\parallel} , α_{\perp} , and f_v . We have used the BUZZARD-V1.1 mock galaxy catalogue to probe these parameter differences by simulating the observation of line-blended galaxies and compared them to the true redshift distributions. We then created a 'shuffled' catalogue; this catalogue uses the same distribution of δ_z values (see equation 5) as the observed catalogue, but with these values randomly shuffled between different galaxies. This results in all correlations between galaxy location and metallicity being erased. By analysing the differences between the parameters fit to these catalogues, we can gain a sense of the potential parameter errors due to the line-blending effect, and to what extent they may require mitigation in order to meet the systematic requirements for *WFIRST*.

We found that errors dependent on the large-scale structure, i.e. the difference between the shuffled and observed catalogues, were $\Delta\alpha_{\parallel} = 0.31 \pm 0.23$ per cent, $\Delta\alpha_{\perp} = -0.10 \pm 0.10$ per cent, and $\Delta f_v = 0.17 \pm 0.33$ per cent ($1.355 \leq z < 1.994$), $\Delta\alpha_{\parallel} = 0.26 \pm 0.17$ per cent, $\Delta\alpha_{\perp} = -0.12 \pm 0.11$ per cent, and $\Delta f_v = -0.20 \pm 0.30$ per cent ($0.705 \leq z < 1.345$), all quoted at the $1.3 \times 10^{-16} \text{ erg cm}^{-2} \text{ s}^{-1}$ flux limit. This uses approximately 18 per cent, 9 per cent, and 7 per cent of their respective error budgets, in an RSS sense. These errors are small – in particular, they are smaller than the upper limits presented by other recent analyses (Faisst et al. 2018) – and can be made smaller still through the use of mitigation techniques described in Section 6. Errors that are dependent on the knowledge of the distribution of galaxy parameters, i.e. the difference between the shuffled and true catalogues, are larger; however these errors are more easily mitigated since the redshift error PDF can be measured by reobserving a small fraction of the sample with high-resolution spectrographs. We estimated that direct mitigation would require reobservation of a few thousand galaxies. The redshift survey C3R2 (Masters et al. 2017) will provide samples of high-resolution spectra in *WFIRST* fields (although the selection criteria are not similar to the *WFIRST* grism survey). We recommend more work to refine this estimate and define the optimal strategy.

It is important to note that this is only a first study of the effect of [N II] + $H\alpha$ line blending on large-scale structure. It examines the effect of one correlation – the mean dependence of [N II]/ $H\alpha$ on stellar mass (at fixed redshift), and hence on environment. Future work should investigate a wider range of astrophysical and instrumental sources of bias on and scatter in the observed redshifts, and treat the possible subtle interactions among all these effects. Specific improvements that would be valuable for the next phase of *WFIRST* studies include:

(i) *Galaxy population*: In this work, we have assumed the environmental dependence is only due to the relationship between galaxy mass and [N II]/ $H\alpha$ ratio. We expect this to be the dominant effect, both because the additional dependence on SFR (at fixed M_* and z) is observed to be weak, and because there is a strong relation between stellar mass and clustering strength. Future work should explore the robustness of these results under different semi-analytic model assumptions, and different prescriptions for the galaxy SEDs

and $H\alpha$ emission line properties. It should also investigate the higher order moments of the $[N II]/H\alpha$ line ratio (e.g. scatter as a function of stellar mass). Realistic correlations of the $[N II]/H\alpha$ line ratio with other properties such as the galaxy radius will also be needed to track the interaction of the $[N II] + H\alpha$ line blending with other instrument-related biases.

(ii) *Instrument and analysis-induced redshift errors*: In this paper, we have *not* included the dependence of the random redshift error on galaxy properties (e.g. size, line flux). Including this is important, but was deferred in this work because it will require significant additional technical steps (particularly in defining the ‘shuffled’ catalogue). In a grism survey, coma or other forms of PSF asymmetry along the dispersion direction result in redshift biases that depend on galaxy properties (size and emission line S/N) as well as field position. While in the real survey these must be addressed with a full survey simulation, a simulation with postage stamps of the emission lines would be a much less computationally demanding intermediate step. This would capture how redshift bias and noise scale with galaxy properties, including dependence on field position and hence the imprint of the tiling strategy (but would not model density-dependent selection effects due to confusion).

(iii) *Survey volume*: In order to investigate higher order effects that contribute to the error budget, larger survey volume is required. This study uses a simulated area of π sr (\sim five times the *WFIRST* reference survey footprint). In this volume it is hard to measure biases that are small compared to *WFIRST* statistical errors, even though the differences of true/observed/shuffled catalogues cancel some of the sampling variance. Ideally we would also have enough simulated realizations to build a mock-based covariance matrix (and thus avoid issues related to variable galaxy number density as a function of redshift and non-Gaussianity of the galaxy density fluctuations).

(iv) *BAO reconstruction*: Reconstruction is a non-linear operation on the survey data, and we should search for possible interaction with density-dependent redshift biases.

In summary, we have presented a simulation-based analysis of the effects of $[N II]$ and $H\alpha$ line blending on BAO and RSD parameter fitting. We conclude that the errors due to the lowest order effect (the trend of $[N II]/H\alpha$ ratio with stellar mass and hence large-scale environment) are small compared to *WFIRST* requirements even without mitigation, and with mitigation should not be a concern for these applications of the *WFIRST* galaxy redshift survey. We have outlined the key improvements needed in future work to study other correlations involving the $[N II]/H\alpha$ ratio and their interaction with instrument and analysis-related redshift errors. We have also concluded that the redshift error probability distribution function will need to be measured accurately; while a brute-force approach seems feasible, we recommend further study of the optimal approach.

ACKNOWLEDGEMENTS

We would like to thank Paul Martini for input on galaxy mass–metallicity relations. We appreciate the many useful conversations with Ami Choi, Niall MacCrann, and Heidi Wu. Further thanks to Paulo Montero-Camacho, Benjamin Buckman, Ben Wibking, and Matthew Digman. We thank Ashley J. Ross for his contributions to our correlation function pipeline. We would also like to thank James Rhoads and Sangeeta Malhotra for providing feedback on our results and taking part in discussions on ideas for future work.

We are also grateful for suggestions from an anonymous referee which improved the paper.

DM, XF, and CMH are supported by the Simons Foundation, the U.S. Department of Energy, the Packard Foundation, the NSF, and NASA. RHW and JD received partial support through NASA contract NNG16PJ25C, from NASA ROSES ATP 16-ATP16-0084 grant, and from the U.S. Department of Energy under contract number DE-AC02-76SF00515. This research used computational resources at SLAC National Accelerator Laboratory, a U.S. Department of Energy Office; the authors thank the support of the SLAC computational team. Many computations in this paper were run on the CCAPP condo of the Ruby Cluster at the Ohio Supercomputer Center (Ohio Supercomputer Center 1987). This research made extensive use of the arXiv and NASA’s Astrophysics Data System for bibliographic information.

REFERENCES

- Alam S. et al., 2015, *ApJS*, 219, 12
Atek H. et al., 2010, *ApJ*, 723, 104
Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
Blake C. et al., 2011, *MNRAS*, 418, 1707
Blanton M. R. et al., 2003, *AJ*, 125, 2348
Blanton M. R. et al., 2017, *AJ*, 154, 28
Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmol. Astropart. Phys.*, 7, 034
Carlson J., Reid B., White M., 2013, *MNRAS*, 429, 1674
Chabrier G., 2003, *PASP*, 115, 763
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, *MNRAS*, 469, 2913
Cohn J. D., 2006, *New A*, 11, 226
Colbert J. W. et al., 2013, *ApJ*, 779, 34
Cole S. et al., 2005, *MNRAS*, 362, 505
Cooper M. C. et al., 2006, *MNRAS*, 370, 198
Cooper M. C., Tremonti C. A., Newman J. A., Zabludoff A. I., 2008, *MNRAS*, 390, 245
Crocce M., Pueblas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
da Cunha E. et al., 2017, *Publ. Astron. Soc. Aust.*, 34, e047
Dawson K. S. et al., 2016, *AJ*, 151, 44
Denicoló G., Terlevich R., Terlevich E., 2002, *MNRAS*, 330, 69
Depagne É., 2015, in Miglio A., Eggenberger P., Girardi L., Montalbán J., eds, *Astrophysics and Space Science Proceedings*, vol. 39, *Asteroseismology of Stellar Populations in the Milky Way*, Springer International Publishing, Cham, Switzerland, p. 147
DeRose J. et al., 2019, preprint (arXiv:1901.02401)
DESI Collaboration, 2016a, preprint (arXiv:1611.00036)
DESI Collaboration, 2016b, preprint (arXiv:1611.00037)
Dressler A. et al., 2012, preprint (arXiv:1210.7809)
Dutton A. A., van den Bosch F. C., Dekel A., 2010, *MNRAS*, 405, 1690
Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
Faisst A. L., Masters D., Wang Y., Merson A., Capak P., Malhotra S., Rhoads J. E., 2018, *ApJ*, 855, 132
Gehrels N., Spergel D., WFIRST SDT Project, 2015, *J. Phys.: Conf. Ser.*, 610, 012007
Gong Q. et al., 2016, in MacEwen H. A., Fazio G. G., Lystrup M., Batalha N., Siegler N., Tong E. C., eds, *Proc. SPIE Conf. Ser.*, Vol. 9904, *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, SPIE, Bellingham, p. 990412
Green J. et al., 2012, preprint (arXiv:1208.4012)
Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, *MNRAS*, 457, 1577
Hirschmann M., Charlot S., Feltre A., Naab T., Choi E., Ostriker J. P., Somerville R. S., 2017, *MNRAS*, 472, 2468
Jones E., Oliphant T., Peterson P. et al., 2001, *SciPy: Open source scientific tools for Python*. <http://www.scipy.org/>
Jones D. H. et al., 2009, *MNRAS*, 399, 683

- Kashino D. et al., 2017, *ApJ*, 835, 88
- Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35
- Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013, *ApJ*, 774, L10
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, 80, 155
- Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, *ApJ*, 772, 119
- Maier C., Ziegler B. L., Lilly S. J., Contini T., Pérez-Montero E., Lamareille F., Bolzonella M., Le Floch E., 2015, *A&A*, 577, A14
- Maiolino R. et al., 2008, *A&A*, 488, 463
- Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, 408, 2115
- Martens D., Hirata C. M., Ross A. J., Fang X., 2018, *MNRAS*, 478, 711
- Masters D. et al., 2014, *ApJ*, 785, 153
- Masters D., Faisst A., Capak P., 2016, *ApJ*, 828, 18
- Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, *ApJ*, 841, 111
- McEwen J. E., Fang X., Hirata C. M., Blazek J. A., 2016, *J. Cosmol. Astropart. Phys.*, 9, 015
- Mehta V. et al., 2015a, *ApJ*, 811, 141
- Mehta V. et al., 2015b, *ApJ*, 811, 141
- Merson A., Wang Y., Benson A., Faisst A., Masters D., Kiessling A., Rhodes J., 2018, *MNRAS*, 474, 177
- Mo H., van den Bosch F. C., White S., 2010, *Galaxy Formation and Evolution*, Cambridge University Press, Cambridge, UK.
- Muzzin A., van Dokkum P., Kriek M., Labbé I., Cury I., Marchesini D., Franx M., 2010, *ApJ*, 725, 742
- Ohio Supercomputer Center, 1987, Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>
- Peebles P. J. E., Hauser M. G., 1974, *ApJS*, 28, 19
- Percival W. J. et al., 2010, *MNRAS*, 401, 2148
- Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, *MNRAS*, 381, 1053
- Perlmutter S. et al., 1999, *ApJ*, 517, 565
- Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, L59
- Pozzetti L. et al., 2016, *A&A*, 590, A3
- Riess A. G. et al., 1998, *AJ*, 116, 1009
- Ross A. J. et al., 2017, *MNRAS*, 464, 1168
- Sánchez A. G., Baugh C. M., Angulo R. E., 2008, *MNRAS*, 390, 1470
- Savaglio S. et al., 2005, *ApJ*, 635, 260
- Seo H.-J., Eisenstein D. J., 2007, *ApJ*, 665, 14
- Shapley A. E. et al., 2015, *ApJ*, 801, 88
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013, *MNRAS*, 428, 1128
- Spergel D. et al., 2015, preprint ([arXiv:1503.03757](https://arxiv.org/abs/1503.03757))
- Springel V., 2005, *MNRAS*, 364, 1105
- Steidel C. C. et al., 2014, *ApJ*, 795, 165
- Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, *ApJ*, 826, 159
- Storey P. J., Zeppen C. J., 2000, *MNRAS*, 312, 813
- Strom A. L., Steidel C. C., Rudie G. C., Trainor R. F., Pettini M., Reddy N. A., 2017, *ApJ*, 836, 164
- Tamura N. et al., 2016, in Evans C. J., Simard L., Takami H., eds, *Proc. SPIE Conf. Ser.*, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99081M
- Villar V., Gallego J., Pérez-González P. G., Pascual S., Noeske K., Koo D. C., Barro G., Zamorano J., 2008, *ApJ*, 677, 169
- Wang L., Reid B., White M., 2014, *MNRAS*, 437, 588
- Wu changeyts E. et al., 2016, *ApJ*, 827, 74
- Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, *MNRAS*, 427, 2146

This paper has been typeset from a \LaTeX file prepared by the author.